

#5

日 本 国 特 許 庁  
JAPAN PATENT OFFICE



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2000年10月23日

出 願 番 号

Application Number:

特願2000-322914

出 願 人

Applicant(s):

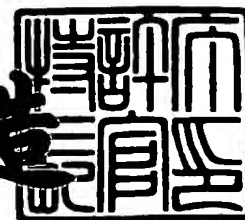
セイコーエプソン株式会社



2001年 9月11日

特許庁長官  
Commissioner,  
Japan Patent Office

及 川 耕 造



出証番号 出証特2001-3083920

【書類名】 特許願

【整理番号】 J0081682

【提出日】 平成12年10月23日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 3/00

【発明者】

【住所又は居所】 長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社内

【氏名】 宮沢 康永

【特許出願人】

【識別番号】 000002369

【氏名又は名称】 セイコーエプソン株式会社

【代表者】 安川 英昭

【代理人】

【識別番号】 100093388

【弁理士】

【氏名又は名称】 鈴木 喜三郎

【連絡先】 0266-52-3139

【選任した代理人】

【識別番号】 100095728

【弁理士】

【氏名又は名称】 上柳 雅誉

【選任した代理人】

【識別番号】 100107261

【弁理士】

【氏名又は名称】 須澤 修

【手数料の表示】

【予納台帳番号】 013044

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9711684

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声認識方法および音声認識処理プログラムを記録した記録媒体ならびに音声認識装置

【特許請求の範囲】

【請求項 1】 種類の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これら雑音が重畳されたそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルを用いることで、雑音の種類に対応する音響モデル群を作成してそれを保持しておく、

認識時には、

雑音が重畳された認識対象音声データに対し、重畳されている雑音の種類を判定し、その判定結果に基づいて、前記雑音の種類に対応した音響モデル群の中から所定の音響モデルを選択するとともに、前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行い、その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行うことを特徴とする音声認識方法。

【請求項 2】 前記雑音除去手法は、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法であって、その場合、前記音響モデル群は、前記種類の異なる雑音が重畳されたそれぞれの音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、雑音除去されたそれぞれの音声データから特徴ベクトルを得て、その特徴ベクトルを用いて作成され、

認識時には、

前記雑音が重畳された認識対象音声データに対し、周波数領域での特徴データを得るための第 1 の音声特徴分析を行い、

その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定された場合には、その分析された特徴データを保存し、音声区間であると判定された場合には、前記保存された特徴データによって、重畳されている雑音

の種類を判定して、その判定結果に基づいて、前記雑音の種類対応に用意された音響モデル群の中から所定の音響モデルを選択し、

前記雑音が重畳された認識対象音声データに対し、前記スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、

その雑音除去された音声データに対し、音声認識に必要な特徴データを得るための第2の音声特徴分析を行い、その特徴分析結果に対し、前記選択された音響モデルを用いて音声認識を行う、

ことを特徴とする請求項1記載の音声認識方法。

【請求項3】 前記雑音除去手法は、ケプストラム平均正規化法による雑音除去手法であって、その場合、前記音響モデルは、前記種類の異なる雑音が重畳されたそれぞれの音声データに対し、ケプストラム平均正規化法による雑音除去手法を用いて雑音除去を行い、それによって得られたそれぞれの音声データの特徴ベクトルを用いて作成され、

認識時には、

前記雑音が重畳された認識対象音声データに対し、ケプストラム係数を表す特徴ベクトルを求めるための第1の音声特徴分析を行い、

その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定された場合には、その特徴ベクトルを保存し、音声区間であると判定された場合には、その音声区間の開始から終了までの音声区間に対応する特徴ベクトルを保存し、前記保存された雑音区間の特徴ベクトルによって、重畳されている雑音の種類を判定して、その判定結果に基づいて、前記雑音の種類対応に用意された音響モデル群の中から所定の音響モデルを選択し、

前記雑音の重畳された音声区間に対し、前記保存されている音声区間の特徴ベクトルを用い、ケプストラム平均正規化法による雑音除去手法を用いて雑音除去処理を行い、

その雑音除去処理後の特徴ベクトルに対し、前記選択された音響モデルを用いて音声認識を行う、

ことを特徴とする請求項1記載の音声認識方法。

【請求項 4】 前記それぞれの雑音対応の音響モデルは、雑音の種類に加え、それぞれの雑音の種類ごとに複数段階の  $S/N$  比にも対応した音響モデルとし、その場合の音響モデルは、複数種類の雑音ごとに  $S/N$  比の異なるそれぞれの雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、このそれぞれ音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルによって作成されたそれぞれの雑音の種類ごとに複数段階の  $S/N$  比に対応した音響モデルであることを特徴とする請求項 1 から 3 のいずれかに記載の音声認識方法。

【請求項 5】 前記音響モデルがそれぞれの雑音の種類ごとに複数段階の  $S/N$  比にも対応した音響モデルである場合、前記雑音が重畳された認識対象音声データに対し、雑音の種類を判定する処理に加え、雑音区間の雑音の大きさと音声区間の音声の大きさから  $S/N$  比を求める処理を行い、判定された雑音種類と求められた  $S/N$  比に基づいて音響モデルの選択を行うことを特徴とする請求項 4 記載の音声認識方法。

【請求項 6】 種類の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これら種類の異なる雑音が重畳されたそれぞれの音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データに対しケプストラム平均正規化法を適用して当該音声区間に対する特徴ベクトルを得て、その特徴ベクトルによって、それぞれの雑音対応の音響モデル群を作成してそれを保存しておき、

認識時には、

雑音の重畳された認識対象音声データに対し、周波数領域での特徴データを得るための第 1 の音声特徴分析を行い、

その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定された場合には、その分析された特徴データを保存し、

音声区間であると判定された場合には、その音声区間に対し前記スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、

その雑音除去された音声区間のデータに対し、ケプストラム係数を求めるための第2の音声特徴分析処理を行い、その音声区間の特徴ベクトルを保存し、

音声区間終了後に、前記保存された雑音区間の特徴データによって、重畳されている雑音の種類を判定して、その判定結果に基づいて、前記雑音対応に用意された音響モデル群の中から所定の音響モデルを選択し、

前記雑音の重畳された音声区間に対し、前記保存されている音声区間の特徴ベクトルに対し、ケプストラム平均正規化法を適用して当該音声区間に対する特徴ベクトルを得て、

そのケプストラム平均正規化法を適用して得られた特徴ベクトルに対し、前記選択された音響モデルを用いて音声認識を行う、

ことを特徴とする音声認識方法。

【請求項7】 前記それぞれの雑音対応の音響モデルは、雑音の種類に加え、それぞれの雑音の種類ごとに複数段階の $S/N$ 比にも対応した音響モデルとし、その場合の音響モデルは、複数種類の雑音ごとに $S/N$ 比の異なるそれぞれの雑音を雑音の種類ごとに重畳させたそれぞれの音声データを作成し、この音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データに対し、ケプストラム平均正規化法を適用して得られたそれぞれの音声データの特徴ベクトルを用いて作成されたそれぞれの雑音の種類ごとに複数段階の $S/N$ 比対応の音響モデルであることを特徴とする請求項6記載の音声認識方法。

【請求項8】 前記音響モデルがそれぞれの雑音の種類ごとに複数段階の $S/N$ 比にも対応した音響モデルである場合、前記雑音が重畳された認識対象音声データに対し、雑音の種類を判定する処理に加え、雑音区間の雑音の大きさと音声区間の音声の大きさから $S/N$ 比を求める処理を行い、判定された雑音種類と求められた $S/N$ 比に基づいて音響モデルの選択を行うことを特徴とする請求項7記載の音声認識方法。

【請求項9】  $S/N$ 比の異なるある特定の種類の雑音がそれぞれの $S/N$ 比ごとに重畳されたそれぞれの音声データを作成し、これらそれぞれの音声デー

タに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルを用いることで、それぞれの S/N 比に対応する音響モデル群を作成してそれを保持しておき、

認識時には、

雑音が重畳された認識対象音声データに対し、重畳されている S/N 比を判定し、その判定結果に基づいて、前記 S/N 比に対応した音響モデル群の中から所定の音響モデルを選択するとともに、前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行い、その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行うことを特徴とする音声認識方法。

【請求項 10】 前記雑音除去手法は、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法であることを特徴とする請求項 9 記載の音声認識方法。

【請求項 11】 前記雑音除去手法は、ケプストラム平均正規化法による雑音除去手法であることを特徴とする請求項 9 記載の音声認識方法。

【請求項 12】 種類の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これら種類の異なる雑音が重畳されたそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データを特徴分析処理して得られた特徴ベクトルによって、雑音の種類に対応する音響モデル群を作成して、それを音響モデル群記憶手段に記憶させる手順と、

雑音が重畳された認識対象音声データに対し、重畳されている雑音の種類を判定し、その判定結果に基づいて、前記音響モデル群記憶手段に記憶された音響モデル群の中から所定の音響モデルを選択する手順と、

前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行う手順と、

その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行う手順と、

をその処理プログラムに含むことを特徴とする音声認識処理プログラムを記録



した記録媒体。

【請求項 1 3】 前記雑音除去手法は、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法であって、その場合、前記音響モデル群は、種類の異なる雑音が重畳されたそれぞれの音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、雑音除去されたそれぞれの音声データから特徴ベクトルを得て、その特徴ベクトルを用いて作成され、

認識時の処理手順は、

前記雑音が重畳された認識対象音声データに対し、周波数領域での特徴データを得るための第 1 の音声特徴分析を行う手順と、

その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定された場合には、その分析された特徴データを保存し、音声区間であると判定された場合には、前記保存された特徴データによって、重畳されている雑音の種類を判定して、その判定結果に基づいて、前記雑音の種類対応に用意された音響モデル群の中から所定の音響モデルを選択する手順と、

前記雑音が重畳された認識対象音声データに対し、前記スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行う手順と、

その雑音除去された音声データに対し、音声認識に必要な特徴データを得るための第 2 の音声特徴分析を行い、その特徴分析結果に対し、前記選択された音響モデルを用いて音声認識を行う手順と、

を含むことを特徴とする請求項 1 2 記載の音声認識処理プログラムを記録した記録媒体。

【請求項 1 4】 前記雑音除去手法は、ケプストラム平均正規化法による雑音除去手法であって、その場合、前記前記音響モデルは、前記種類の異なる雑音が重畳されたそれぞれの音声データに対し、ケプストラム平均正規化法による雑音除去手法を用いて雑音除去を行い、それによって得られたそれぞれの音声データの特徴ベクトルを用いて作成され、

認識時の処理手順は、

前記雑音が重畳された認識対象音声データに対し、ケプストラム係数を表す特徴ベクトルを求めるための第 1 の音声特徴分析を行う手順と、

その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定された場合には、その特徴ベクトルを保存し、音声区間であると判定された場合には、その音声区間の開始から終了までの音声区間に対応する特徴ベクトルを保存し、前記保存された雑音区間の特徴ベクトルによって、重畳されている雑音の種類を判定して、その判定結果に基づいて、前記雑音の種類対応に用意された音響モデル群の中から所定の音響モデルを選択する手順と、

前記雑音の重畳された音声区間に対し、前記保存されている音声区間の特徴ベクトルを用い、ケプストラム平均正規化法による雑音除去手法を用いて雑音除去処理を行う手順と、

その雑音除去処理後の特徴ベクトルに対し、前記選択された音響モデルを用いて音声認識を行う手順と、

を含むことを特徴とする請求項 1 2 記載の音声認識処理プログラムを記録した記録媒体。

【請求項 1 5】 前記それぞれの雑音対応の音響モデルは、雑音の種類に加え、それぞれの雑音の種類ごとに複数段階の  $S/N$  比にも対応した音響モデルとし、その場合の音響モデルは、複数種類の雑音ごとに  $S/N$  比の異なるそれぞれの雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、このそれぞれ音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルによって作成されたそれぞれの雑音の種類ごとに複数段階の  $S/N$  比に対応した音響モデルであることを特徴とする請求項 1 2 から 1 4 のいずれかに記載の音声認識処理プログラムを記録した記録媒体。

【請求項 1 6】 前記音響モデルがそれぞれの雑音の種類ごとに複数段階の  $S/N$  比にも対応した音響モデルである場合、前記雑音が重畳された認識対象音声データに対し、雑音の種類を判定する処理に加え、雑音区間の雑音の大きさと音声区間の音声の大きさから  $S/N$  比を求める処理を行い、判定された雑音種類

と求められた  $S/N$  比に基づいて音響モデルの選択を行うことを特徴とする請求項 15 記載の音声認識処理プログラムを記録した記録媒体。

【請求項 17】 種類の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これら種類の異なる雑音が重畳されたそれぞれの音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データに対しケプストラム平均正規化法を適用して当該音声区間に対する特徴ベクトルを得て、その特徴ベクトルによって、それぞれの雑音対応の音響モデル群を作成し、それを音響モデル群記憶手段に記憶させる手順と、

雑音の重畳された認識対象音声データに対し、周波数領域での特徴データを得るための第 1 の音声特徴分析を行う手順と、

その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定された場合には、その分析された特徴データを保存する手順と、

音声区間であると判定された場合には、その音声区間に対し前記スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行う手順と、

その雑音除去された音声区間のデータに対し、ケプストラム係数を求めるための第 2 の音声特徴分析処理を行い、その音声区間の特徴ベクトルを保存する手順と、

音声区間終了後に、前記保存された雑音区間の特徴分析データによって、重畳されている雑音の種類を判定して、その判定結果に基づいて、前記雑音対応に用意された音響モデル群の中から所定の音響モデルを選択する手順と、

前記雑音の重畳された音声区間に対し、前記保存されている音声区間の特徴ベクトルに対し、ケプストラム平均正規化法を適用して当該音声区間に対する特徴ベクトルを得る手順と、

そのケプストラム平均正規化法を適用して得られた特徴ベクトルに対し、前記選択された音響モデルを用いて音声認識を行う手順と、

をその処理プログラムに含むことを特徴とする音声認識処理プログラムを記録した記録媒体。

【請求項 1 8】 前記それぞれの雑音対応の音響モデルは、雑音の種類に加え、それぞれの雑音の種類ごとに複数段階の  $S/N$  比にも対応した音響モデルとし、その場合の音響モデルは、複数種類の雑音ごとに  $S/N$  比の異なるそれぞれの雑音を雑音の種類ごとに重畳させたそれぞれの音声データを作成し、この音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データに対し、ケプストラム平均正規化法を適用して得られたそれぞれの音声データの特徴ベクトルを用いて作成されたそれぞれの雑音の種類ごとに複数段階の  $S/N$  比対応の音響モデルであることを特徴とする請求項 1 7 記載の音声認識処理プログラムを記録した記録媒体。

【請求項 1 9】 前記音響モデルがそれぞれの雑音の種類ごとに複数段階の  $S/N$  比にも対応した音響モデルである場合、前記雑音が重畳された認識対象音声データに対し、雑音の種類を判定する処理に加え、雑音区間の雑音の大きさと音声区間の音声の大きさから  $S/N$  比を求める処理を行い、判定された雑音種類と求められた  $S/N$  比に基づいて音響モデルの選択を行うことを特徴とする請求項 1 8 記載の音声認識処理プログラムを記録した記録媒体。

【請求項 2 0】  $S/N$  比の異なるある特定の種類の雑音がそれぞれの  $S/N$  比ごとに重畳されたそれぞれの音声データを作成し、これらそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルを用いることで、それぞれの  $S/N$  比に対応する音響モデル群を作成して、それを音響モデル群記憶手段に記憶させる手順と、

雑音が重畳された認識対象音声データに対し、重畳されている  $S/N$  比を判定し、その判定結果に基づいて、前記  $S/N$  比に対応した音響モデル群の中から所定の音響モデルを選択する手順と、

前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行う手順と、

その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行う手順と、

を処理プログラムに含むことを特徴とする音声認識処理プログラムを記録した記録媒体。

【請求項 2 1】 前記雑音除去手法は、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法であることを特徴とする請求項 2 0 記載の音声認識処理プログラムを記録した記録媒体。

【請求項 2 2】 前記雑音除去手法は、ケプストラム平均正規化法による雑音除去手法であることを特徴とする請求項 2 0 記載の音声認識処理プログラムを記録した記録媒体。

【請求項 2 3】 種類の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これら種類の異なる雑音が重畳されたそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルによって作成された雑音の種類に対応する音響モデル群と、

この音響モデル群を記憶する音響モデル群記憶手段と、

雑音が重畳された認識対象音声データに対し、重畳されている雑音の種類を判定する雑音判定手段と、

その判定結果に基づいて、前記雑音の種類に対応した音響モデル群の中から所定の音響モデルを選択する音響モデル群選択手段と、

前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行う雑音除去手段と、

その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行う音声認識手段と、

を有したことを特徴とする音声認識装置。

【請求項 2 4】 前記雑音除去手段が行う雑音除去手法は、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法であって、その場合、前記音響モデル群は、前記種類の異なる雑音が重畳されたそれぞれの音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、雑音除去されたそれぞれの音声データから特徴ベクトルを得て、その特徴

ベクトルを用いて作成され、

これによって作成された音響モデル群を記憶する音響モデル群記憶手段と、

前記雑音が重畳された認識対象音声データに対し、周波数領域での特徴分析データを得るための第 1 の音声特徴分析を行う第 1 の音声特徴分析手段と、

その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定した場合には、その雑音区間の特徴データを特徴データ記憶手段に保存する雑音区間／音声区間判定手段と、

音声区間であると判定された場合には、前記保存された特徴データによって、重畳されている雑音の種類を判定する雑音種類判定手段と、

その判定結果に基づいて、前記雑音の種類対応に用意された前記音響モデル群の中から所定の音響モデルを選択する音響モデル選択手段と、

前記雑音が重畳された認識対象音声データに対し、前記スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行う雑音除去手段と、その雑音除去された音声データに対し、音声認識に必要な特徴データを得るための第 2 の音声特徴分析を行う第 2 の音声特徴分析手段と、

その特徴分析結果に対し、前記選択された音響モデルを用いて音声認識を行う音声認識手段と、

を有したことを特徴とする請求項 2 3 記載の音声認識装置。

【請求項 2 5】 前記雑音除去手段が行う雑音除去手法は、ケプストラム平均正規化法による雑音除去手法であって、その場合、前記音響モデルは、前記種類の異なる雑音が重畳されたそれぞれの音声データに対し、ケプストラム平均正規化法による雑音除去手法を用いて雑音除去を行い、それによって得られたそれぞれの音声データの特徴ベクトルを用いて作成され、

これによって作成された音響モデル群を記憶する音響モデル群記憶手段と、

前記雑音が重畳された認識対象音声データに対し、ケプストラム係数を表す特徴ベクトルを求めるための特徴分析を行う特徴分析手段と、

その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定した場合には、その雑音区間の特徴ベクトルを特徴データ記憶手段に保存

し、音声区間であると判定した場合には、その音声区間の特徴ベクトルを特徴分析データ記憶手段に保存する雑音区間／音声区間判定手段と、

この特徴データ記憶手段に保存された雑音区間の特徴ベクトルによって、重畳されている雑音の種類を判定する雑音種類判定手段と、

その判定結果に基づいて、前記雑音の種類対応に用意された前記音響モデル群の中から所定の音響モデルを選択する音響モデル選択手段と、

前記雑音の重畳された音声区間に対し、前記保存されている音声区間の特徴ベクトルを用い、ケプストラム平均正規化法による雑音除去手法を用いて雑音除去処理を行う雑音除去手段と、

その雑音除去処理によって得られた特徴ベクトルに対し、前記選択された音響モデルを用いて音声認識を行う音声認識手段と、

有することを特徴とする請求項 2 3 記載の音声認識装置。

【請求項 2 6】 前記それぞれの雑音対応の音響モデルは、雑音の種類に加え、それぞれの雑音の種類ごとに複数段階の  $S/N$  比にも対応した音響モデルとし、その場合の音響モデルは、複数種類の雑音ごとに  $S/N$  比の異なるそれぞれの雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、このそれぞれ音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルによって作成されたそれぞれの雑音の種類ごとに複数段階の  $S/N$  比に対応した音響モデルであることを特徴とする請求項 2 3 から 2 5 のいずれかに記載の音声認識装置。

【請求項 2 7】 前記音響モデルがそれぞれの雑音の種類ごとに複数段階の  $S/N$  比にも対応した音響モデルである場合、前記雑音種類判定手段は、前記雑音が重畳された認識対象音声データに対し、雑音の種類を判定する処理に加え、雑音区間の雑音の大きさと音声区間の音声の大きさから  $S/N$  比を求める処理を行い、前記音響モデル選択部は、判定された雑音種類と求められた  $S/N$  比に基づいて音響モデルの選択を行うことを特徴とする請求項 2 6 記載の音声認識装置。

【請求項 2 8】 種類の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これら種類の異なる雑音が重畳されたそれぞれの音声デ

ータに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データに対しケプストラム平均正規化法を適用して当該音声区間に対する特徴ベクトルを得て、その特徴ベクトルによって作成されたそれぞれの雑音対応の音響モデル群と、

この音響モデル群を記憶する音響モデル群記憶手段と、

雑音の重畳された認識対象音声データに対し、周波数領域での特徴データを得るための第 1 の音声特徴分析を行う第 1 の音声特徴分析手段と、

その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定された場合には、その特徴データを特徴データ記憶手段に保存する雑音区間／音声区間判定手段と、

音声区間であると判定された場合には、その音声区間に対し前記スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法によるノイズ除去手法を用いて雑音除去を行う雑音除去手段と、

その雑音除去された音声区間のデータに対し、ケプストラム係数を求めるための第 2 の特徴分析処理を行いその音声区間の特徴ベクトルを特徴データ記憶手段に保存する第 2 の音声特徴分析部と、

音声区間終了後に、前記保存された雑音区間の特徴データによって、重畳されている雑音の種類を判定する雑音種類判定手段と、

その判定結果に基づいて、前記雑音対応に用意された音響モデルの中から所定の音響モデルを選択する音響モデル選択手段と、

前記雑音の重畳された音声区間に対し、前記保存されている音声区間の特徴ベクトルを用い、ケプストラム平均正規化法を適用して当該音声区間に対する特徴ベクトルを出力するケプストラム平均正規化演算部と、

その特徴ベクトルに対し、前記選択された音響モデルを用いて音声認識を行う音声認識部と、

を有することを特徴とする音声認識装置。

【請求項 29】 前記それぞれの雑音対応の音響モデルは、雑音の種類に加え、それぞれの雑音の種類ごとに複数段階の S/N 比にも対応した音響モデルと



し、その場合の音響モデルは、複数種類の雑音ごとに  $S/N$  比の異なるそれぞれの雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、それぞれ音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データに対し、ケプストラム平均正規化法を適用して得られたそれぞれの音声データの特徴ベクトルを用いて作成されたそれぞれの雑音の種類ごとに複数段階の  $S/N$  比対応の音響モデルであることを特徴とする請求項 28 記載の音声認識装置。

【請求項 30】 前記音響モデルがそれぞれの雑音の種類ごとに複数段階の  $S/N$  比にも対応した音響モデルである場合、前記雑音種類判定手段は、前記雑音が重畳された認識対象音声データに対し、雑音の種類を判定する処理に加え、雑音区間の雑音の大きさと音声区間の音声の大きさから  $S/N$  比を求める処理を行い、前記音響モデル選択部は、判定された雑音種類と求められた  $S/N$  比に基づいて音響モデルの選択を行うことを特徴とする請求項 29 記載の音声認識装置。

【請求項 31】  $S/N$  比の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これらそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルによって作成された  $S/N$  比に対応する音響モデル群と、

この音響モデル群を記憶する音響モデル群記憶手段と、

雑音が重畳された認識対象音声データに対し、重畳されている雑音の  $S/N$  比を判定する  $S/N$  比判定手段と、

その判定結果に基づいて、前記  $S/N$  比に対応した音響モデル群の中から所定の音響モデルを選択する音響モデル群選択手段と、

前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行う雑音除去手段と、

その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行う音声認識手段と、

を有したことを特徴とする音声認識装置。

【請求項 3 2】 前記雑音除去手法は、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法であることを特徴とする請求項 3 1 記載の音声認識装置。

【請求項 3 3】 前記雑音除去手法は、ケプストラム平均正規化法による雑音除去手法であることを特徴とする請求項 3 1 記載の音声認識装置。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は種々の背景雑音が存在する環境下においても高い認識性能を可能とする音声認識方法および音声認識処理プログラムを記録した記録媒体ならびに音声認識装置に関する。

【0 0 0 2】

【従来の技術】

近年、音声認識機能を搭載した機器が広く用いられるようになってきている。このような機器の使用環境は様々であり、雑音の多い環境下で使用せざるを得ない場合も多い。

【0 0 0 3】

このような場合、当然のことながら雑音に対する対策を講じる必要が出てくる。雑音の一例としては、たとえば、自動車の走行音、エアコンディショナ（エアコンという）の運転音などの定常的な雑音が身近なものとして考えられるが、これらの定常的な雑音の存在する環境下での音声認識を行う方法として、従来、以下に示すような音声認識方法が用いられている。

【0 0 0 4】

その第 1 の例として、雑音のない環境下で収録した音声データに上述したような定常的な雑音から得られた雑音データを重畳させた音声データを生成し、この生成された音声データを用いて学習された音声認識用の音響モデルを作成し、その音響モデルを用いて音声認識を行う方法がある。

【0 0 0 5】

また、第 2 の例としては、スペクトラル・サブトラクション（Spectral Subtr

action) などの雑音除去方法を用いて音声認識を行う方法もある。この音声認識方法は、入力音声データから雑音成分を除去して、雑音の除去された音声データに対して音声認識を行うが、その場合でも、上述同様、雑音のない環境下で収録した音声データに定常的な雑音から得られた雑音データを重畳させた音声データを生成し、この生成された音声データからスペクトラル・サブトラクション法による雑音除去方法を用いて雑音を除去したのちに、その雑音の除去された音声データを用いて学習した音声認識用の音響モデルを作成しておき、その音響モデルを用いて音声認識を行うことがなされている。

## 【0006】

## 【発明が解決しようとする課題】

上述したような音声認識方法を採用することによって、何の対策も講じない場合に比べ、雑音環境下における認識性能の向上はある程度は可能となると考えられるが、まだまだ問題点も多い。

## 【0007】

すなわち、定常的な雑音は、上述したような自動車の走行音、エアコンの運転音などの他にも、雑踏による雑音など様々な種類があり、それぞれが異なった性質を持っている。

## 【0008】

上述した従来の2つの例で述べた音響モデルは、音響モデルを学習する際、ある特定の雑音のみを用いて学習している場合が多い。たとえば、自動車の走行音を雑音データとして用い、その雑音データを音声データに重畳させた音声データを生成し、この生成された音声データからスペクトラル・サブトラクションによる雑音除去方法を用いて雑音を除去したのちに、その雑音の除去された音声データを用いて学習した音声認識用の音響モデルを作成する。

## 【0009】

このように、ある特定の雑音について作成された音響モデルを用いて音声認識を行えば、そのような雑音が存在する環境下では比較的好結果が得られるが、状況によっては、それ以外の雑音が存在することも十分あり得る。その場合の認識率は当然のことながら低下する。

## 【 0 0 1 0 】

また、雑音の種類以外にも、実際に認識すべき音声信号と雑音信号の比である S/N 比の大きさによっても認識性能は異なってくる。

## 【 0 0 1 1 】

そこで本発明は、雑音の種類や S/N 比の大きさに対応した高い認識性能を得ることができ、しかも、演算能力の低い CPU を用いた安価なハードウェアで実現可能とすることを目的としている。

## 【 0 0 1 2 】

## 【課題を解決するための手段】

上述した目的を達成するために本発明の音声認識方法は、種類の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これら雑音が重畳されたそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルを用いることで、雑音の種類に対応する音響モデル群を作成してそれを保持しておく。そして、認識時には、雑音が重畳された認識対象音声データに対し、重畳されている雑音の種類を判定し、その判定結果に基づいて、前記雑音の種類に対応した音響モデル群の中から所定の音響モデルを選択するとともに、前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行い、その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行うようにしている。

## 【 0 0 1 3 】

また、本発明の音声認識処理プログラムを記録した記録媒体は、種類の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これら種類の異なる雑音が重畳されたそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データを特徴分析処理して得られた特徴ベクトルによって、雑音の種類に対応する音響モデル群を作成して、それを音響モデル群記憶手段に記憶させる手順と、雑音が重畳された認識対象音声データに対し、重畳されている雑音の種類を判定し、その判定結果に基づいて、前記音響モデル群記憶手段に記憶された音響モデル群の中から所定の

音響モデルを選択する手順と、前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行う手順と、その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行う手順とをその処理プログラムに含むものである。

## 【 0 0 1 4 】

これら各発明において、前記雑音除去手法は、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法であって、その場合、前記音響モデル群は、前記種類の異なる雑音が重畳されたそれぞれの音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、雑音除去されたそれぞれの音声データから特徴ベクトルを得て、その特徴ベクトルを用いて作成されている。そして、認識時には、前記雑音が重畳された認識対象音声データに対し、周波数領域での特徴データを得るための第1の音声特徴分析を行い、その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定された場合には、その分析された特徴データを保存し、音声区間であると判定された場合には、前記保存された特徴データによって、重畳されている雑音の種類を判定して、その判定結果に基づいて、前記雑音の種類対応に用意された音響モデル群の中から所定の音響モデルを選択し、前記雑音が重畳された認識対象音声データに対し、前記スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去された音声データに対し、音声認識に必要な特徴データを得るための第2の音声特徴分析を行い、その特徴分析結果に対し、前記選択された音響モデルを用いて音声認識を行うようにしている。

## 【 0 0 1 5 】

また、前記雑音除去手法は、ケプストラム平均正規化法による雑音除去手法であってもよく、その場合、前記音響モデルは、前記種類の異なる雑音が重畳されたそれぞれの音声データに対し、ケプストラム平均正規化法による雑音除去手法を用いて雑音除去を行い、それによって得られたそれぞれの音声データの特徴ベクトルを用いて作成されている。そして、認識時には、前記雑音が重畳された認

識対象音声データに対し、ケプストラム係数を表す特徴ベクトルを求めるための第1の音声特徴分析を行い、その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定された場合には、その特徴ベクトルを保存し、音声区間であると判定された場合には、その音声区間の開始から終了までの音声区間に対応する特徴ベクトルを保存し、前記保存された雑音区間の特徴ベクトルによって、重畳されている雑音の種類を判定して、その判定結果に基づいて、前記雑音の種類対応に用意された音響モデル群の中から所定の音響モデルを選択し、前記雑音の重畳された音声区間に対し、前記保存されている音声区間の特徴ベクトルを用い、ケプストラム平均正規化法による雑音除去手法を用いて雑音除去処理を行い、その雑音除去処理後の特徴ベクトルに対し、前記選択された音響モデルを用いて音声認識を行うようにしている。

## 【0016】

さらに、前記それぞれの雑音対応の音響モデルは、雑音の種類に加え、それぞれの雑音の種類ごとに複数段階の $S/N$ 比にも対応した音響モデルとすることも可能で、その場合の音響モデルは、複数種類の雑音ごとに $S/N$ 比の異なるそれぞれの雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、このそれぞれ音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルによって作成されたそれぞれの雑音の種類ごとに複数段階の $S/N$ 比に対応した音響モデルとしている。

## 【0017】

そして、音響モデルがそれぞれの雑音の種類ごとに複数段階の $S/N$ 比にも対応した音響モデルである場合、前記雑音が重畳された認識対象音声データに対し、雑音の種類を判定する処理に加え、雑音区間の雑音の大きさと音声区間の音声の大きさから $S/N$ 比を求める処理を行い、判定された雑音種類と求められた $S/N$ 比に基づいて音響モデルの選択を行うようにしている。

## 【0018】

また、本発明のもう一つの音声認識方法は、種類の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これら種類の異なる雑音が重畳されたそれぞれの音声データに対し、スペクトラル・サブトラクション法または

連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データに対しケプストラム平均正規化法を適用して当該音声区間に対する特徴ベクトルを得て、その特徴ベクトルによって、それぞれの雑音対応の音響モデル群を作成してそれを保存しておく。そして、認識時には、雑音の重畳された認識対象音声データに対し、周波数領域での特徴データを得るための第 1 の音声特徴分析を行い、その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定された場合には、その分析された特徴データを保存し、音声区間であると判定された場合には、その音声区間に対し前記スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去された音声区間のデータに対し、ケプストラム係数を求めるための第 2 の音声特徴分析処理を行い、その音声区間の特徴ベクトルを保存し、音声区間終了後に、前記保存された雑音区間の特徴データによって、重畳されている雑音の種類を判定して、その判定結果に基づいて、前記雑音対応に用意された音響モデル群の中から所定の音響モデルを選択し、前記雑音の重畳された音声区間に対し、前記保存されている音声区間の特徴ベクトルに対し、ケプストラム平均正規化法を適用して当該音声区間に対する特徴ベクトルを得て、そのケプストラム平均正規化法を適用して得られた特徴ベクトルに対し、前記選択された音響モデルを用いて音声認識を行うようにしている。

## 【 0 0 1 9 】

また、もう一つの音声認識処理プログラムを記録した記録媒体は、種類の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これら種類の異なる雑音が重畳されたそれぞれの音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データに対しケプストラム平均正規化法を適用して当該音声区間に対する特徴ベクトルを得て、その特徴ベクトルによって、それぞれの雑音対応の音響モデル群を作成し、それを音響モデル群記憶手段に記憶させる手順と、 雑音の重畳された認識対象音声データに対し、周波数領域での特徴データを得るための第 1 の音声特徴分析を行

う手順と、その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定された場合には、その分析された特徴データを保存する手順と、音声区間であると判定された場合には、その音声区間に対し前記スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行う手順と、その雑音除去された音声区間のデータに対し、ケプストラム係数を求めるための第2の音声特徴分析処理を行い、その音声区間の特徴ベクトルを保存する手順と、音声区間終了後に、前記保存された雑音区間の特徴分析データによって、重畳されている雑音の種類を判定して、その判定結果に基づいて、前記雑音対応に用意された音響モデル群の中から所定の音響モデルを選択する手順と、前記雑音の重畳された音声区間に対し、前記保存されている音声区間の特徴ベクトルに対し、ケプストラム平均正規化法を適用して当該音声区間に対する特徴ベクトルを得る手順と、そのケプストラム平均正規化法を適用して得られた特徴ベクトルに対し、前記選択された音響モデルを用いて音声認識を行う手順とをその処理プログラムに含むものである。

## 【0020】

これら音声認識方法および音声認識処理プログラムを記録した記録媒体において、前記それぞれの雑音対応の音響モデルは、雑音の種類に加え、それぞれの雑音の種類ごとに複数段階のS/N比にも対応した音響モデルとし、その場合の音響モデルは、複数種類の雑音ごとにS/N比の異なるそれぞれの雑音を雑音の種類ごとに重畳させたそれぞれの音声データを作成し、この音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データに対し、ケプストラム平均正規化法を適用して得られたそれぞれの音声データの特徴ベクトルを用いて作成されたそれぞれの雑音の種類ごとに複数段階のS/N比対応の音響モデルとしている。

## 【0021】

そして、音響モデルがそれぞれの雑音の種類ごとに複数段階のS/N比にも対応した音響モデルである場合、前記雑音が重畳された認識対象音声データに対し、雑音の種類を判定する処理に加え、雑音区間の雑音の大きさと音声区間の音声



の大きさから  $S/N$  比を求める処理を行い、判定された雑音種類と求められた  $S/N$  比に基づいて音響モデルの選択を行うようにする。

## 【 0 0 2 2 】

さらに、本発明の音声認識方法は、 $S/N$  比の異なるある特定の種類の雑音がそれぞれの  $S/N$  比ごとに重畳されたそれぞれの音声データを作成し、これらそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルを用いることで、それぞれの  $S/N$  比に対応する音響モデル群を作成してそれを保持しておき、認識時には、雑音が重畳された認識対象音声データに対し、重畳されている  $S/N$  比を判定し、その判定結果に基づいて、前記  $S/N$  比に対応した音響モデル群の中から所定の音響モデルを選択するとともに、前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行い、その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行うようにしたものでもよい。

## 【 0 0 2 3 】

さらに、本発明の音声認識処理プログラムを記録した記録媒体は、 $S/N$  比の異なるある特定の種類の雑音がそれぞれの  $S/N$  比ごとに重畳されたそれぞれの音声データを作成し、これらそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルを用いることで、それぞれの  $S/N$  比に対応する音響モデル群を作成して、それを音響モデル群記憶手段に記憶させる手順と、雑音が重畳された認識対象音声データに対し、重畳されている  $S/N$  比を判定し、その判定結果に基づいて、前記  $S/N$  比に対応した音響モデル群の中から所定の音響モデルを選択する手順と、前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行う手順と、その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行う手順とを含んだ処理プログラムとしてもよい。

## 【 0 0 2 4 】

これら各発明において、雑音除去手法は、スペクトラル・サブトラクション法

または連続スペクトラル・サブトラクション法であってもよく、ケプストラム平均正規化法による雑音除去手法であってもよい。

## 【 0 0 2 5 】

また、本発明の音声認識装置は、種類の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これら種類の異なる雑音が重畳されたそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルによって作成された雑音の種類に対応する音響モデル群と、この音響モデル群を記憶する音響モデル群記憶手段と、雑音が重畳された認識対象音声データに対し、重畳されている雑音の種類を判定する雑音判定手段と、その判定結果に基づいて、前記雑音の種類に対応した音響モデル群の中から所定の音響モデルを選択する音響モデル群選択手段と、

前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行う雑音除去手段と、その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行う音声認識手段とを有した構成としている。

## 【 0 0 2 6 】

この音声認識装置において、前記雑音除去手段が行う雑音除去手法は、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法であって、その場合、前記音響モデル群は、前記種類の異なる雑音が重畳されたそれぞれの音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、雑音除去されたそれぞれの音声データから特徴ベクトルを得て、その特徴ベクトルを用いて作成される。そして、これによって作成された音響モデル群を記憶する音響モデル群記憶手段と、前記雑音が重畳された認識対象音声データに対し、周波数領域での特徴分析データを得るための第1の音声特徴分析を行う第1の音声特徴分析手段と、その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定した場合には、その雑音区間の特徴データを特徴データ記憶手段に保存する雑音区間／音声区間判定手段と、音声区間であると判定された場合には、前記保存された特徴データによって、重畳されてい

る雑音の種類を判定する雑音種類判定手段と、その判定結果に基づいて、前記雑音の種類対応に用意された前記音響モデル群の中から所定の音響モデルを選択する音響モデル選択手段と、前記雑音が重畳された認識対象音声データに対し、前記スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行う雑音除去手段と、その雑音除去された音声データに対し、音声認識に必要な特徴データを取得するための第2の音声特徴分析を行う第2の音声特徴分析手段と、その特徴分析結果に対し、前記選択された音響モデルを用いて音声認識を行う音声認識手段とを有した構成としている。

## 【0027】

また、前記雑音除去手段が行う雑音除去手法は、ケプストラム平均正規化法による雑音除去手法であって、その場合、前記音響モデルは、前記種類の異なる雑音が重畳されたそれぞれの音声データに対し、ケプストラム平均正規化法による雑音除去手法を用いて雑音除去を行い、それによって得られたそれぞれの音声データの特徴ベクトルを用いて作成される。そして、これによって作成された音響モデル群を記憶する音響モデル群記憶手段と、前記雑音が重畳された認識対象音声データに対し、ケプストラム係数を表す特徴ベクトルを求めるための特徴分析を行う特徴分析手段と、その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定した場合には、その雑音区間の特徴ベクトルを特徴データ記憶手段に保存し、音声区間であると判定した場合には、その音声区間の特徴ベクトルを特徴分析データ記憶手段に保存する雑音区間／音声区間判定手段と、この特徴データ記憶手段に保存された雑音区間の特徴ベクトルによって、重畳されている雑音の種類を判定する雑音種類判定手段と、その判定結果に基づいて、前記雑音の種類対応に用意された前記音響モデル群の中から所定の音響モデルを選択する音響モデル選択手段と、前記雑音の重畳された音声区間に対し、前記保存されている音声区間の特徴ベクトルを用い、ケプストラム平均正規化法による雑音除去手法を用いて雑音除去処理を行う雑音除去手段と、その雑音除去処理によって得られた特徴ベクトルに対し、前記選択された音響モデルを用いて音声認識を行う音声認識手段とを有した構成としている。

## 【 0 0 2 8 】

前記それぞれの雑音対応の音響モデルは、雑音の種類に加え、それぞれの雑音の種類ごとに複数段階の  $S/N$  比にも対応した音響モデルとし、その場合の音響モデルは、複数種類の雑音ごとに  $S/N$  比の異なるそれぞれの雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、このそれぞれ音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルによって作成されたそれぞれの雑音の種類ごとに複数段階の  $S/N$  比に対応した音響モデルである。

## 【 0 0 2 9 】

そして、音響モデルがそれぞれの雑音の種類ごとに複数段階の  $S/N$  比にも対応した音響モデルである場合、前記雑音種類判定手段は、前記雑音が重畳された認識対象音声データに対し、雑音の種類を判定する処理に加え、雑音区間の雑音の大きさと音声区間の音声の大きさから  $S/N$  比を求める処理を行い、前記音響モデル選択部は、判定された雑音種類と求められた  $S/N$  比に基づいて音響モデルの選択を行うようにしている。

## 【 0 0 3 0 】

また、本発明のもう一つの音声認識装置は、種類の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これら種類の異なる雑音が重畳されたそれぞれの音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データに対しケプストラム平均正規化法を適用して当該音声区間に対する特徴ベクトルを得て、その特徴ベクトルによって作成されたそれぞれの雑音対応の音響モデル群と、この音響モデル群を記憶する音響モデル群記憶手段と、雑音の重畳された認識対象音声データに対し、周波数領域での特徴データを得るための第 1 の音声特徴分析を行う第 1 の音声特徴分析手段と、その特徴分析結果を用いて、雑音区間か音声区間かを判定し、雑音区間であると判定された場合には、その特徴データを特徴データ記憶手段に保存する雑音区間／音声区間判定手段と、音声区間であると判定された場合には、その音声区間に対し前記スペクトラル・サブトラクション法または連続スペクトラ

ル・サブトラクション法によるノイズ除去手法を用いて雑音除去を行う雑音除去手段と、その雑音除去された音声区間のデータに対し、ケプストラム係数を求めるための第2の特徴分析処理を行いその音声区間の特徴ベクトルを特徴データ記憶手段に保存する第2の音声特徴分析部と、音声区間終了後に、前記保存された雑音区間の特徴データによって、重畳されている雑音の種類を判定する雑音種類判定手段と、その判定結果に基づいて、前記雑音対応に用意された音響モデルの中から所定の音響モデルを選択する音響モデル選択手段と、前記雑音の重畳された音声区間に対し、前記保存されている音声区間の特徴ベクトルを用い、ケプストラム平均正規化法を適用して当該音声区間に対する特徴ベクトルを出力するケプストラム平均正規化演算部と、その特徴ベクトルに対し、前記選択された音響モデルを用いて音声認識を行う音声認識部とを有した構成としている。

## 【0031】

このような音声認識装置において、前記それぞれの雑音対応の音響モデルは、雑音の種類に加え、それぞれの雑音の種類ごとに複数段階の $S/N$ 比にも対応した音響モデルとし、その場合の音響モデルは、複数種類の雑音ごとに $S/N$ 比の異なるそれぞれの雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、それぞれ音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データに対し、ケプストラム平均正規化法を適用して得られたそれぞれの音声データの特徴ベクトルを用いて作成されたそれぞれの雑音の種類ごとに複数段階の $S/N$ 比対応の音響モデルである。

## 【0032】

そして、音響モデルがそれぞれの雑音の種類ごとに複数段階の $S/N$ 比にも対応した音響モデルである場合、前記雑音種類判定手段は、前記雑音が重畳された認識対象音声データに対し、雑音の種類を判定する処理に加え、雑音区間の雑音の大きさと音声区間の音声の大きさから $S/N$ 比を求める処理を行い、前記音響モデル選択部は、判定された雑音種類と求められた $S/N$ 比に基づいて音響モデルの選択を行うようにしている。

## 【0033】

さらに本発明の音声認識装置は、 $S/N$ 比の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これらそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルによって作成された $S/N$ 比に対応する音響モデル群と、この音響モデル群を記憶する音響モデル群記憶手段と、雑音が重畳された認識対象音声データに対し、重畳されている雑音の $S/N$ 比を判定する $S/N$ 比判定手段と、その判定結果に基づいて、前記 $S/N$ 比に対応した音響モデル群の中から所定の音響モデルを選択する音響モデル群選択手段と、前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行う雑音除去手段と、その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行う音声認識手段とを有した構成としたものであってもよい。

## 【 0 0 3 4 】

その場合の雑音除去手法は、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法であってもよく、ケプストラム平均正規化法による雑音除去手法であってもよい。

## 【 0 0 3 5 】

このように本発明は、種類の異なる雑音が重畳されたそれぞれの音声データを作成し、これらそれぞれの雑音が重畳されたそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データを用いて、雑音の種類に対応する音響モデルを作成しておく。そして、実際の認識時には、雑音が重畳された認識対象音声データに対し、重畳されている雑音の種類を判定して、その判定結果に基づいて、前記雑音の種類に対応した音響モデルの中から所定の音響モデルを選択するとともに、前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行い、その雑音除去された音声データに対し、前記選択された音響モデルを用いて音声認識を行うようにしている。

## 【 0 0 3 6 】

これによって、重畳されている雑音の種類に応じた最適な音響モデルを用いて

の認識処理が可能となり、所定の雑音の存在する環境下であっても高い認識率を得ることができる。

## 【 0 0 3 7 】

特に、機器の使用環境に 2、3 種類の雑音が定常的に存在するような場合、それらの雑音ごとの音響モデルを作成し、その音響モデルを用いて、上述したような音声認識処理を行うことで、高い認識率を実現できる。

## 【 0 0 3 8 】

そして、本発明で用いられる雑音除去手法の 1 つとしては、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法であって、その場合、前記音響モデル作成時における雑音除去は、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法を用いて行う。また、実際の認識時には、雑音区間の特徴分析データによって、重畳されている雑音の種類を判定したのち、その判定結果に基づいて、最適な音響モデルを選択するとともに、前記雑音が重畳された認識対象音声データに対し、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去された音声データを特徴分析して得られた結果に対し、前記選択された音響モデルを用いて音声認識を行うようにしている。

## 【 0 0 3 9 】

このように、雑音除去方法としてスペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法を用いることにより、雑音除去処理を少ない演算量で行うことができ、演算能力の低い CPU でも十分対応することができる。これにより、小規模で安価なハードウェア上での実現が可能となる。また、このスペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法は自動車の走行音やエアコンの運転音、街中の雑踏などの雑音（一般に加法性雑音といわれている）の除去に効果があるとされているので、このような雑音の多い環境下で用いられることが多い機器に適用されることで大きな効果が得られる。

## 【 0 0 4 0 】

また、雑音除去手法の他の例として、ケプストラム平均正規化法による雑音除去手法を用いることもこともできる。その場合、前記音響モデル作成時における雑音除去は、ケプストラム平均正規化法を用いて行う。また、実際の認識時には、雑音区間の特徴分析データによって、重畳されている雑音の種類を判定したのち、その判定結果に基づいて、最適な音響モデルを選択するとともに、前記雑音が重畳された認識対象音声データの音声区間に対し、ケプストラム平均正規化法を用いて雑音除去処理を行い、その雑音除去処理によって得られた特徴ベクトルに対し、前記選択された音響モデルを用いて音声認識を行うようにしている。

## 【0041】

このように、雑音除去方法としてケプストラム平均正規化法を用いることにより、上述同様、雑音除去処理を少ない演算量で行うことができ、演算能力の低いCPUでも十分対応することができる。これにより、小規模で安価なハードウェア上での実現が可能となる。また、このケプストラム平均正規化法はマイクロホンの特性やエコーなど空間伝達特性に由来する歪みなどの雑音（一般に乗法性雑音といわれている）の除去に効果があるとされているので、このような雑音が発生しやすい環境下で用いられることが多い機器に適用されることで大きな効果が得られる。

## 【0042】

さらに、それぞれの雑音対応の音響モデルは、雑音の種類に加え、それぞれの雑音の種類ごとに複数段階のS/N比にも対応した音響モデルとし、実際の認識時には、前記雑音が重畳された認識対象音声データに対し、雑音区間の雑音の大きさと音声区間の音声の大きさからS/N比を求め、求められたS/N比と雑音の種類に応じた音響モデルを選択するようにしているので、雑音の種類だけではなくその大きさに応じた最適な音響モデルを用いての認識が行える。これによって、それぞれの雑音環境下において音声認識を行う際、より一層、高い認識率を得ることが可能となる。

## 【0043】

また、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法とケプストラム平均正規化法の両方を用いた音響モデルを作成する



ことも可能である。この場合、実際の音声認識を行う場合も、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法を用いた雑音除去を行ったのちに、その雑音除去された音声データに対し、ケプストラム平均正規化法で特徴ベクトルを生成し、それを音声認識用の特徴ベクトルとして音声認識部に渡すようにしているので、さらに高い認識性能を得ることができ、また、この場合、前述した加法性雑音や乗法性雑音など幅広い雑音に対する対応が可能となる。

## 【0044】

さらに本発明は、ある特定の決まった雑音について複数の $S/N$ を考慮した音声認識を行うことも可能である。その場合、 $S/N$ 比の異なるある特定の種類の雑音がそれぞれの $S/N$ 比ごとに重畳されたそれぞれの音声データを作成し、これらそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルを用いることで、それぞれの $S/N$ 比に対応する音響モデル群を作成しておく。そして、実際の認識時には、雑音が重畳された認識対象音声データに対し、重畳されている $S/N$ 比を判定し、その判定結果に基づいて、それぞれの $S/N$ 比に対応した音響モデル群の中から所定の音響モデルを選択し、前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行い、その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行うようにしている。

## 【0045】

これは、雑音の種類は特定できてもその大きさ( $S/N$ 比)が変動することの多い環境下での音声認識を行う場合に好都合なものとなり、そのような環境下での認識率を高くすることができる。

## 【0046】

## 【発明の実施の形態】

以下、本発明の実施の形態について説明する。なお、この実施の形態で説明する内容は、本発明の音声認識方法および音声認識装置についての説明であるとともに、本発明の音声認識処理プログラムを記録した記録媒体における音声認識処

理プログラムの具体的な処理内容をも含むものである。

【0047】

本発明は基本的には、処理対象となる音声に重畳している雑音を除去して、雑音の除去された音声データに対して音声認識を行うものであるが、その音声認識に用いられる音響モデルは、雑音（定常的な雑音）の種類を幾つか想定し、それぞれの雑音をある音声に対する音声データ（雑音の全く無いクリーンな音声データ）に重畳させて雑音の重畳された音声データを生成し、その雑音の重畳された音声データから雑音を除去する処理を行い、その雑音除去処理後の音声波形（雑音の無いクリーンな音声データとは多少異なる）を用いて音響モデルを作成する。

【0048】

すなわち、予め用意された幾つかの雑音の種類ごとに、上述した手順でノイズの種類ごとにその雑音の除去された音響モデルが作成されることになる。

【0049】

そして、実際の音声認識を行う場合には、認識対象の音声データに重畳されている雑音の種類を判定するとともに、その雑音を除去する処理を行い、雑音の種類に応じて音響モデルを選択して、選択された音響モデルを用いて音声認識処理を行う。

【0050】

さらに、これらそれぞれの雑音の種類とともに、雑音と音声データの大きさの比である  $S/N$  比を幾つかに設定した音響モデルを作成する。たとえば、雑音の種類を雑音  $N1$ 、雑音  $N2$ 、雑音  $N3$  の3種類を選んだとすれば、これら雑音の種類だけを考慮した場合には、3種類の音響モデルが作成されるが、それぞれの雑音について2段階の  $S/N$  比を考慮するとすれば、それぞれの雑音について雑音の大きさを2種類設定して、上述した処理を行って音響モデルを作成することになるので、作成される音響モデルは6種類となる。

【0051】

たとえば、 $S/N$  比がある値  $L1$  未満 ( $S/N < L1$ ) の場合と、 $L1$  以上 ( $S/N \geq L1$ ) の場合の2段階に設定したとすれば、雑音  $N1$  に対しては、 $S$

／N比がL1未満の場合の音響モデルと、L1以上の場合の音響モデルの2つの音響モデルが作成される。同様に、雑音N2、N3に対しても、それぞれ、S／NがL1未満の場合の音響モデルと、L1以上の場合の音響モデルの2個ずつの音響モデルが作成され、合計6種類の音響モデルが作成されることになる。

#### 【0052】

ところで、上述の雑音除去を行う技術としては、前述したように、スペクトラル・サブトラクション (Spectral Subtraction: 以下、SSという) 法または連続スペクトラル・サブトラクション (Continuous Spectral Subtraction: 以下、CSSという) があるが、これは、特に、自動車の走行音、エアコンの運転音、街の雑踏などどこに音源が存在するのかが特定しにくい雑音 (前述したように、加法性雑音と呼ばれている) の除去に効果のある方法といわれている。

#### 【0053】

これらSS法またはCSS法とは別に、ケプストラム平均正規化 (Cepstrum Mean Normalization: 以下、CMNという) 法による雑音除去方法もある。この方法は、マイクロホン特性やエコーなど空間伝達特性に由来する歪みなどの雑音 (前述したように、乗法性雑音と呼ばれている) の除去に効果がある方法であるといわれている。

#### 【0054】

そこで本発明の実施の形態では、雑音除去方法としてSS法またはCSS法を用いた場合を第1の実施の形態、CMN法を用いた場合を第2の実施の形態、その両方を用いた場合を第3の実施の形態として説明する。

#### 【0055】

##### 〔第1の実施の形態〕

図1はこの第1の実施の形態の音声認識装置の概略構成を示す図であり、構成要素のみを列挙すれば、マイクロホン1、アンプやA／D変換器を有する入力音声処理部2、第1の音声特徴分析部3、雑音区間／音声区間判定部4、特徴分析データ記憶部5、雑音種類判定／音響モデル選択部6、音響モデル群記憶部7、雑音除去部8、第2の音声特徴分析部9、音声認識部10、言語モデル記憶部11などを有した構成となっている。これら各構成要素の機能などについては図2

のフローチャートを参照した動作説明により逐次説明する。

【0056】

図2において、A/D変換後の認識対象音声データに対し、まず、第1の音声特徴分析部3によって、1フレームごと（1フレームの時間長はたとえば20数msec程度）に音声特徴分析が行われる（ステップs1）。この音声特徴分析は、周波数領域での音声特徴分析であり、その周波数分析手法として、たとえば、FFT（高速フーリエ変換）などを用いた音声特徴分析であるとする。

【0057】

そして、雑音区間／音声区間判定部4は、その音声特徴分析結果から得られるパワーの大きさや周波数の特徴などから、音声データが雑音区間であるか音声区間であるかを判定する（ステップs2）。その判定結果により雑音区間であると判定された場合には、最新のnフレーム分の特徴データを特徴データ記憶部5に記憶させておく（ステップs3）。このステップs1～s3の処理を繰り返し、やがて、音声区間に入ったと判定されると、雑音種類判定／音響モデル選択部6により雑音種類判定動作と音響モデル選択動作に入る。この雑音種類判定動作と音響モデル選択動作について以下に説明する。

【0058】

まず、この雑音種類判定動作と音響モデル選択動作の開始指示があるか否かを見て（ステップs4）、開始指示があれば、雑音の種類と大きさ（S/N比）を判定し、かつ、その判定結果に基づく音響モデル選択動作を行う（ステップs5）。

【0059】

ここで、雑音の種類と大きさの判定は、ステップs3において特徴データ記憶部5に記憶された最新のnフレーム分の雑音区間の特徴データおよび第1の音声特徴分析処理で得られる音声区間の幾つかのフレームごとの特徴データを用いて行う。これらそれぞれの特徴データからは周波数成分の特徴の他にパワーなども得られるため、雑音の種類やパワーがわかるとともに、音声のパワーがわかる。

【0060】

たとえば、この第1の実施の形態では、雑音として自動車の走行音、エアコン

の運転音、街中の雑踏などの定常的な雑音を想定している。ここでは、このような定常的な雑音として3種類を考え、それを雑音N1、雑音N2、雑音N3で表すものとする。したがって、雑音区間のnフレーム分の特徴データを調べることによって、それが雑音N1、N2、N3のどれに近いかを判定することができる。

#### 【0061】

また、雑音のパワーと音声のパワーがわかれば、S/N比を求めることができる。なお、S/N比を求めるには、音声区間のパワーがある程度の大きさを持ったところでS/N比を計算する必要があるため、たとえば、音声区間における数フレーム分もしくは全フレーム分の最大値や平均値を用いて、S/N比の計算を行う。

#### 【0062】

このようにして、雑音の種類が判定されるとともにS/N比が求められると、次に、音響モデル選択動作を行う。この第1の実施の形態では、音響モデルは、3種類の定常的な雑音N1、N2、N3を想定し、これら3種類の雑音N1、N2、N3に対し、S/N比の値がL1未満の音響モデルと、L1以上の音響モデルを用意してある。

#### 【0063】

たとえば、この第1の実施の形態では、雑音の種類が雑音N1でS/N比がL1未満である場合には音響モデルM1、雑音N1でS/N比がL1以上である場合には音響モデルM2、雑音N2でS/N比がL1未満である場合には音響モデルM3、雑音N2でS/N比がL1以上である場合には音響モデルM4、雑音N3でS/N比がL1未満である場合には音響モデルM5、雑音N3でS/N比がL1以上である場合には音響モデルM6というように対応付けられているとする。したがって、音響モデル群記憶部7には、これら6種類の音響モデルM1、M2、・・・、M6が保存されている。これらの音響モデルM1、M2、・・・、M6は次のようにして作成される。

#### 【0064】

すなわち、雑音N1、N2、N3とそれぞれの雑音について2段階のS/N比

( $L1$  未満か  $L1$  以上か) を有する 6 パターンの雑音を用意し、これら 6 パターンの雑音を雑音の全くない音声データに重畳させることで、6 パターンの音声データを作成する。

## 【0065】

この 6 パターンの音声データは、 $S/N$  比が  $L1$  未満の雑音  $N1$  が重畳された音声データ、 $S/N$  比が  $L1$  以上の雑音  $N1$  が重畳された音声データ、 $S/N$  比が  $L1$  未満の雑音  $N2$  が重畳された音声データ、 $S/N$  比が  $L1$  以上の雑音  $N2$  が重畳された音声データ、 $S/N$  比が  $L1$  未満の雑音  $N3$  が重畳された音声データ、 $S/N$  比が  $L1$  以上の雑音  $N3$  が重畳された音声データの 6 パターンの音声データである。

## 【0066】

これら 6 パターンのそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去された 6 パターンの音声データを特徴分析処理して得られた特徴ベクトルを用いることで、6 種類の音響モデル  $M1$ ,  $M2$ , ...,  $M6$  が作成される。

## 【0067】

ここで、ステップ  $s5$  における処理において、雑音の種類が雑音  $N1$  に近いと判定され、求められた  $S/N$  比が  $(S/N) < L1$ 、つまり  $L1$  未満であった場合には、音響モデル群記憶部 7 から音響モデル  $M1$  が選択される。

## 【0068】

このようにして、ノイズの種類と  $S/N$  比に応じた音響モデルが選択されると、次に、雑音除去部 8 による雑音除去処理がなされる (ステップ  $s6$ )。この雑音除去処理は、この第 1 の実施の形態では  $SS$  法または  $CSS$  法による雑音除去処理であり、前述したステップ  $s3$  において特徴データ記憶部 5 に記憶された最新の  $n$  フレーム分の雑音区間の特徴データと、音声区間の特徴データを用いてスペクトラル減算を行う。これによって、雑音の除去された音声データを得ることができる。ただし、このような雑音除去処理されたあとであっても、音声データには、雑音の影響がわずかではあるが残されたものとなっている。

## 【0069】

そして、その雑音除去処理後の音声データに対して第2の音声特徴分析部9が特徴分析処理を行う（ステップs7）。この第2の音声特徴分析部9による特徴分析処理をここでは第2の音声特徴分析処理と呼ぶことにする。

#### 【0070】

この第2の音声特徴分析処理は、音声認識部10が音声認識を行う際に用いるケプストラム係数を求める処理である。なお、ステップs1にて行われている特徴分析処理がたとえばFFTなどを用いた周波数分析手法であり、その特徴分析結果が周波数領域での音声特徴データとなっているので、この第2の音声特徴分析処理では、ケプストラム係数としてメルケプストラム係数（Mel Frequency Cepstrum Coefficients）を求める。

#### 【0071】

この第2の音声特徴分析処理によって得られたメルケプストラム係数は音声認識部10に与えられ、音声認識部10では、そのメルケプストラム係数に対して音声認識処理を行うが、このとき用いる音響モデルは、ステップs5によって選択された音響モデル（上述した例では音響モデルM1）であり、その音響モデルM1と言語モデル記憶部11に保存されている言語モデルを用いて音声認識を行う。

#### 【0072】

また、ステップs7における第2の音声特徴分析のあとは、音声区間終了か否かを判断して（ステップs8）、音声区間がすべて終了していれば、処理は終了し、音声区間が終わっていないければ、ステップs1に処理が戻って同様の処理を行う。

#### 【0073】

すなわち、第1の音声特徴分析を行い（ステップs1）、雑音区間か音声区間かを判定し（ステップs2）、この判定結果が音声区間であれば、ステップs4以降の処理に入るが、このとき、音響モデル選択動作開始指示がない場合には、雑音の種類と大きさ（S/N比）の判定およびその判定結果に基づく音響モデル選択動作が終了しているか否かを判断して（ステップs9）、その処理が終了していれば、雑音除去処理（ステップs6）を行い、処理が終了していないければ、

第1の音声特徴分析処理によって得られた音声区間の特徴データを記憶する処理を行う（ステップs10）。

## 【0074】

以上のような一連の処理が音声区間が終了するまで行われる。以上説明したように、音声認識対象となる音声データに対し、その音声データに重畳している雑音の種類とS/N比の大きさに応じた音響モデルが選択され、選択された音響モデルと予め用意されている言語モデルを用いて音声認識を行うようにしている。

## 【0075】

なお、この第1の実施の形態で用いられる6種類の音響モデルM1, M2, ..., M6は、前述したように、2段階のS/N比を有する3種類の雑音N1, N2, N3を、音声データ（雑音の全く無いクリーンな音声データ）に重畳させて、雑音の重畳された6パターンの音声データを生成し、その6パターンの音声データに対しそれぞれ雑音を除去する処理（SS法またはCSS法による雑音除去処理）を行い、その雑音除去処理後の6パターンの音声データ（雑音の無いクリーンな音声データとは異なり、ノイズの影響が多少残された音声データ）を用いて作成されたものである。つまり、これら6種類の音響モデルは、実際の音声認識処理対象となる音声データに近い音声データにより作成された音響モデルであると言える。

## 【0076】

したがって、実際の音声認識処理対象となる音声データに対し、その音声データに重畳されている雑音の種類とS/N比の大きさに基づいて、最適な音響モデルが選択され、選択された音響モデルを用いて音声認識を行うことにより、より一層、高い認識性能を得ることができる。

## 【0077】

また、この第1の実施の形態では、雑音除去方法として、SS法またはCSS法を用いているので、雑音除去処理を少ない演算量で行うことができ、演算能力の低いCPUでも十分対応することができる。

## 【0078】

これにより、小規模で安価なハードウェア上での実現が可能となる。また、こ



のSS法は自動車の走行音やエアコンの運転音、街中の雑踏などの雑音除去に効果があるとされているので、このような雑音の多い環境下で用いられることが多い機器に適用されることで大きな効果が得られる。

## 【 0 0 7 9 】

## 〔第2の実施の形態〕

第2の実施の形態は、雑音除去方法として、ケプストラム平均正規化法（CMN法）を用いたものであり、図3はこの第2の実施の形態の音声認識装置の概略構成を示す図であり、構成要素のみを列挙すれば、マイクロホン1、アンプやA/D変換器を有する入力音声処理部2、音声特徴分析部21、雑音区間／音声区間判定部4、特徴データ記憶部5、雑音種類判定／音響モデル選択部6、音響モデル群記憶部7、雑音除去部8、音声認識部10、言語モデル記憶部11などを有した構成となっている。これら各構成要素の機能などについては図4のフローチャートを参照した動作説明により逐次説明する。

## 【 0 0 8 0 】

図4において、まず、音声特徴分析部21がA/D変換後の処理対象音声データに対し、1フレームごと（1フレームの時間長はたとえば20数msec程度）に音声特徴分析を行う（ステップs21）。この音声特徴分析はこの第2の実施の形態ではケプストラム係数（たとえば、メルケプストラム係数やLPCケプストラム係数）を求めるための特徴分析であるとする。

## 【 0 0 8 1 】

そして、その音声特徴分析結果に基づいて、雑音区間であるか音声区間であるかを雑音区間／音声区間判定部4によって判定し（ステップs22）、雑音区間であると判定された場合には、この雑音区間／音声区間判定部4は、さらに、その雑音区間が音声区間の時間軸方向前方に存在する雑音区間であるか、音声区間の時間軸方向後方に存在する雑音区間であるかを判定する（ステップs23）。

## 【 0 0 8 2 】

この判定の結果、音声区間の時間軸方向前方に存在する雑音区間である場合には、特徴分析されて得られた最新のn1フレーム分の特徴データ（ケプストラム係数の特徴ベクトル）を特徴データ記憶部5に記憶させる（ステップs24）。

## 【 0 0 8 3 】

また、雑音区間であるか音声区間であるかを判定した結果が、音声区間であると判定された場合には、その音声区間（音声区間の開始から終了まで）を構成する  $n 2$  フレーム分の特徴データ（ケプストラム係数の特徴ベクトル）を特徴データ記憶部 5 に記憶する（ステップ  $s 2 5$ ）。

## 【 0 0 8 4 】

さらに音声特徴分析を繰り返し、雑音区間であるか音声区間であるかを判定した結果が、雑音区間であると判定され、かつ、その雑音区間が音声区間の時間軸方向後方に存在する雑音区間であると判定された場合には（ステップ  $s 2 1$ 、 $s 2 2$ 、 $s 2 3$ ）、音声区間が終了したものとして、音声区間終了後の  $n 3$  フレーム分の特徴データ（ケプストラム係数の特徴ベクトル）を特徴データ記憶部 5 に記憶する（ステップ  $s 2 6$ ）。

## 【 0 0 8 5 】

そして、この  $n 3$  フレーム分の記憶処理が終了したか否かを判断して（ステップ  $s 2 7$ ）、処理が終了していれば、雑音種類判定／音響モデル選択部 6 により雑音種類判定動作と音響モデル選択動作に入る（ステップ  $s 2 8$ ）。この雑音種類判定動作と音響モデル選択動作について以下に説明する。

## 【 0 0 8 6 】

この雑音の種類と大きさ（ $S/N$  比）の判定および音響モデル選択動作は、それまでに特徴データ記憶部 5 に記憶されている  $n 1$ 、 $n 2$  フレーム分のそれぞれの特徴データを用いて行う。

## 【 0 0 8 7 】

すなわち、雑音の種類は、雑音区間の特徴データ（たとえば  $n 1$  フレーム分の特徴データ）を用いて、雑音がどの雑音に近いかを判定することができ、 $S/N$  比は雑音区間を特徴分析することによって得られるパワーの大きさと音声区間のパワーの大きさによって求めることができる。

## 【 0 0 8 8 】

なお、この第 2 の実施の形態においても、3 種類の雑音  $N 1$ 、 $N 2$ 、 $N 3$  に対応した処理を行うものとする。

## 【 0 0 8 9 】

そして、これら雑音の種類の評定結果と、求められた  $S/N$  比の大きさに基づいて、どの音響モデルを用いるかの音響モデル選択動作を行う。この音響モデル選択動作は、前述の第 1 の実施の形態同様、たとえば、雑音の種類が雑音  $N_1$  に近いと判定され、かつ、 $S/N$  比が  $L_1$  未満であった場合には、音響モデル  $M_1$  が選択されるといった動作である。

## 【 0 0 9 0 】

なお、この第 2 の実施の形態においても、第 1 の実施の形態同様、雑音の種類と  $S/N$  比の大きさに応じて 6 個の音響モデル  $M_1$ ,  $M_2$ ,  $\dots$ ,  $M_6$  が用意されるものとする。

## 【 0 0 9 1 】

すなわち、この第 2 の実施の形態も第 1 の実施の形態同様、雑音  $N_1$  で  $S/N$  比が  $L_1$  未満である場合には音響モデル  $M_1$ 、雑音  $N_1$  で  $S/N$  比が  $L_1$  以上である場合には音響モデル  $M_2$ 、雑音  $N_2$  で  $S/N$  比が  $L_1$  未満である場合には音響モデル  $M_3$ 、雑音  $N_2$  で  $S/N$  比が  $L_1$  以上である場合には音響モデル  $M_4$ 、雑音  $N_3$  で  $S/N$  比が  $L_1$  未満である場合には音響モデル  $M_5$ 、雑音  $N_3$  で  $S/N$  比が  $L_1$  以上である場合には音響モデル  $M_6$  というように対応付けられているとする。したがって、音響モデル群記憶部 7 には、これら 6 種類の音響モデル  $M_1$ ,  $M_2$ ,  $\dots$ ,  $M_6$  が保存されている。

## 【 0 0 9 2 】

なお、この第 2 の実施の形態においては、CMN（ケプストラム平均正規化法）による雑音除去法を用いているので、音響モデル  $M_1$ ,  $M_2$ ,  $\dots$ ,  $M_6$  は CMN 法を用いて作成されたものである。これらの音響モデル  $M_1$ ,  $M_2$ ,  $\dots$ ,  $M_6$  は次のようにして作成される。

## 【 0 0 9 3 】

すなわち、雑音  $N_1$ ,  $N_2$ ,  $N_3$  とそれぞれの雑音について 2 段階の  $S/N$  比（ $L_1$  未満か  $L_1$  以上か）を有する 6 パターンの雑音を用意し、これら 6 パターンの雑音を雑音の全くない音声データに重畳させることで、6 パターンの音声データを作成する。

## 【 0 0 9 4 】

この6パターンの音声データは、 $S/N$ 比が $L1$ 未満の雑音 $N1$ が重畳された音声データ、 $S/N$ 比が $L1$ 以上の雑音 $N1$ が重畳された音声データ、 $S/N$ 比が $L1$ 未満の雑音 $N2$ が重畳された音声データ、 $S/N$ 比が $L1$ 以上の雑音 $N2$ が重畳された音声データ、 $S/N$ 比が $L1$ 未満の雑音 $N3$ が重畳された音声データ、 $S/N$ 比が $L1$ 以上の雑音 $N3$ が重畳された音声データの6パターンの音声データである。

## 【 0 0 9 5 】

これら6パターンのそれぞれの音声データに対し、CMN法による雑音除去手法を用いて雑音除去を行い、その雑音除去された6パターンの音声データの特徴ベクトルを用いることで、6種類の音響モデル $M1$ 、 $M2$ 、 $\dots$ 、 $M6$ が作成される。

## 【 0 0 9 6 】

ここで、ステップ $s28$ における処理において、雑音の種類が雑音 $N1$ に近いと判定され、求められた $S/N$ 比が $L1$ 未満であった場合には、音響モデル群記憶部7から音響モデル $M1$ が選択される。

## 【 0 0 9 7 】

ところで、この雑音の種類と大きさ( $S/N$ 比)の判定動作を行う場合、 $n1$ フレーム分の特徴データ(音声区間の前方に存在する雑音の特徴データ)と、 $n2$ フレーム分の特徴データ(音声区間の開始から終了までの特徴データ)だけを用いてもそれらを判定することができるが、 $n3$ フレーム分の特徴データ(音声区間の後方に存在する雑音の特徴データ)をも用いるようにしてもよい。

## 【 0 0 9 8 】

そして次に、雑音除去部8がCMN法を用いた雑音除去処理を行うが、このCMN法による雑音除去処理は、まず、音声区間の音声特徴分析結果による特徴ベクトル( $n2$ フレーム分の特徴ベクトル)を用い、その $n2$ フレーム分の平均の特徴ベクトルを求める(ステップ $s29$ )。

## 【 0 0 9 9 】

なお、この平均の特徴ベクトルを求める際、 $n2$ フレーム分の特徴ベクトルだ

けを用いるのではなく、 $n_1$ 、 $n_2$ 、 $n_3$ の全ての特徴ベクトルを用いて求めるようにすることもできるが、ここでは、音声区間の開始から終了までを構成する  $n_2$  フレーム分のみの特徴データを用いて行うものとする。

## 【0100】

たとえば、 $n_2 = 20$  とすれば、20 フレーム分の特徴ベクトル（これを  $C_1$ 、 $C_2$ 、 $\dots$ 、 $C_{20}$  で表し、これら各特徴ベクトル  $C_1$ 、 $C_2$ 、 $\dots$ 、 $C_{20}$  は、それぞれ、たとえば10次元の成分を有している）の平均を求める。求められた平均の特徴ベクトルを  $C_m$  とする。

## 【0101】

次に、求められた平均の特徴ベクトルを用い、音声区間（ここでは20フレーム分）の特徴ベクトルを再計算する（ステップs30）。この再計算というのは、音声区間を構成する20フレーム分のそれぞれのフレームごとの特徴ベクトル  $C_1$ 、 $C_2$ 、 $\dots$ 、 $C_{20}$  から、平均の特徴ベクトル  $C_m$  を引き算するもので、この例では、 $C_1' = C_1 - C_m$ 、 $C_2' = C_2 - C_m$ 、 $\dots$ 、 $C_{20}' = C_{20} - C_m$  を行う。そして、求められた  $C_1'$ 、 $C_2'$ 、 $\dots$ 、 $C_{20}'$  が雑音除去処理後の20フレーム分の特徴ベクトルとなる。

## 【0102】

この特徴ベクトル  $C_1'$ 、 $C_2'$ 、 $\dots$ 、 $C_{20}'$  が音声認識部10に与えられ、音声認識部10では、選択された音響モデルと予め用意されている言語モデル11を用いた音声認識処理を行う。

## 【0103】

このように、第2の実施の形態においても前述した第1の実施の形態と同様、雑音の種類とS/N比の大きさに応じた音響モデルが選択され、選択された音響モデルと言語モデル記憶部11に保存されている言語モデルを用いて音声認識を行うようにしている。

## 【0104】

なお、この第2の実施の形態で用いられる6種類の音響モデルは、第1の実施の形態同様、2段階のS/N比を有する3種類の雑音  $N_1$ 、 $N_2$ 、 $N_3$  を音声データ（雑音の全く無いクリーンな音声データ）に重畳させて、雑音の重畳された

6 パターンの音声データを生成し、その 6 パターンの音声データに対しそれぞれ雑音を除去する処理（CMN 法による雑音除去処理）を行い、その雑音除去処理後の 6 パターンの音声データ（雑音の無いクリーンな音声データとは異なり、雑音の影響が多少残された音声データ）を用いて作成されたものである。つまり、実際の音声認識処理対象となる音声データに近い音声データにより作成された音響モデルであると言える。

## 【 0 1 0 5 】

したがって、実際の音声認識処理対象となる音声データに対し、その音声データに重畳されている雑音の種類と  $S/N$  比の大きさに基づいて、最適な音響モデルを選択し、選択された音響モデルを用いて音声認識を行うことにより、より一層、高い認識性能を得ることができる。

## 【 0 1 0 6 】

また、この第 2 の実施の形態の雑音除去法としての CMN 法は、少ない演算量で雑音除去を行うことができ、演算能力の低い CPU でも十分対応することができ、小規模で安価なハードウェア上での実現が可能となる。また、この CMN 法はマイクロホンの特性やエコーなど空間伝達特性に由来する雑音（乗法性雑音）の除去に効果があるとされているので、このような雑音が発生しやすい環境下で用いられることが多い機器に適用されることで大きな効果が得られる。

## 【 0 1 0 7 】

## 〔第 3 の実施の形態〕

この第 3 の実施の形態は、第 1 の実施の形態と第 2 の実施の形態を組み合わせたものである。この第 3 の実施の形態においても、第 1 および第 2 の実施の形態同様、雑音の種類と  $S/N$  比の大きさに応じて 6 個の音響モデル  $M_1$ ,  $M_2$ ,  $\dots$ ,  $M_6$  が用意されているものとするが、この第 3 の実施の形態において用いられる音響モデルは、以下のようにして作成される。

## 【 0 1 0 8 】

前述したように、2 段階の  $S/N$  比を有する 3 種類の雑音  $N_1$ ,  $N_2$ ,  $N_3$  を音声データ（雑音の全く無いクリーンな音声データ）に重畳させて、雑音の重畳された 6 パターンの音声データを生成し、その 6 パターンの音声データに対しそ

れぞれ雑音を除去する処理（S S法またはC S S法による雑音除去処理）を行い、その雑音除去処理後の6パターンの音声データ（雑音の無いクリーンな音声データとは異なり、雑音の影響が多少残された音声データ）を生成する。

#### 【0109】

そして、このS S法またはC S S法により雑音除去された6パターンの音声データのそれぞれの音声区間に対しCMN法を適用する。すなわち、前述したように、それぞれの音声データにおける音声区間を特徴分析して得られた特徴ベクトル（ $n_2$ フレーム分の特徴ベクトル）を用い、その $n_2$ フレーム分の平均の特徴ベクトルを求める。たとえば、 $n_2 = 20$ とすれば、20フレーム分の特徴ベクトル（これを $C_1, C_2, \dots, C_{20}$ で表し、これら各特徴ベクトル $C_1, C_2, \dots, C_{20}$ は、それぞれ、たとえば10次元の成分を有している）の平均 $C_m$ とする。

#### 【0110】

次に、求められた平均の特徴ベクトルを用い、音声区間（ここでは20フレーム分）の特徴ベクトルを再計算、つまり、 $C_1' = C_1 - C_m, C_2' = C_2 - C_m, \dots, C_{20}' = C_{20} - C_m$ を行い、求められた $C_1', C_1', \dots, C_{20}'$ を20フレーム分（音声区間分）のそれぞれのフレームごとの特徴ベクトルとし、これらそれぞれのフレームごとの特徴ベクトルを用いて音響モデルを作成する。

#### 【0111】

このような処理を、3種類のノイズ $N_1, N_2, N_3$ ごとにそれぞれ2種類の $S/N$ 比の大きさを設定して行うことで、6個の音響モデル $M_1, M_2, \dots, M_6$ が作成される。

#### 【0112】

図5はこの第3の実施の形態の音声認識装置の概略構成を示す図であり、構成要素のみを列挙すれば、マイクロホン1、アンプやA/D変換器を有する入力音声処理部2、第1の音声特徴分析部3、雑音区間／音声区間判定部4、特徴データ記憶部5、雑音種類判定／音響モデル選択部6、音響モデル群記憶部7、雑音除去部8、第2の音声特徴分析部9、CMN演算部（CMN法による雑音除

去部) 31、音声認識部10、言語モデル記憶部11などを有した構成となっている。以下、図6のフローチャートを参照して説明する。

### 【0113】

図6において、まず、第1の音声特徴分析部3によって、A/D変換後の認識対象音声データに対し、1フレームごと(1フレームの時間長はたとえば20数msec程度)に音声特徴分析が行われる(ステップs41)。この音声特徴分析は、周波数領域での音声特徴分析であり、ここでは前述同様、FFT(高速フーリエ変換)などを用いた周波数分析手法を用いるものとする。

### 【0114】

その音声特徴分析結果に基づいて、雑音区間であるか音声区間であるかを雑音区間/音声区間判定部4によって判定し(ステップs42)、雑音区間であると判定された場合には、雑音区間/音声区間判定部4は、さらに、その雑音区間が音声区間の時間軸方向前方に存在する雑音区間であるか、音声区間の時間軸方向後方に存在する雑音区間であるかを判定する(ステップs43)。そして、音声区間の時間軸方向前方に存在する雑音区間である場合には、最新のn1フレーム分の特徴データを特徴データ記憶部5に記憶させる(ステップs44)。

### 【0115】

また、雑音区間であるか音声区間であるかを判定した結果が、音声区間であると判定された場合には、雑音除去部8によってSS法またはCSS法による雑音除去処理を行う(ステップs45)。そして、その雑音除去処理後の音声データに対し、第2の特徴分析部9が特徴分析処理を行い(ステップs46)、それによって得られた音声特徴データ(特徴ベクトル)を記憶させておく(ステップs47)。なお、この第2の音声特徴分析処理はメルケプストラム係数を求めるための特徴分析処理である。

### 【0116】

そして、ステップs41に処理が戻り、第1の音声特徴分析処理が繰り返され、その特徴分析結果に基づいて、雑音区間であるか音声区間であるかを判定し、その結果が、雑音区間であると判定され、かつ、その雑音区間が音声区間の時間軸方向後方に存在する雑音区間であると判定された場合には(ステップs41、



s 4 2, s 4 3)、音声区間終了と判断して、ステップ s 4 8 の雑音種類判定動作と音響モデル選択動作処理に入る。

#### 【0 1 1 7】

この雑音の種類と大きさ ( $S/N$  比) の判定および音響モデル選択動作は、それまでに記憶されている  $n 1$  フレーム分および  $n 2$  フレーム分のそれぞれの音声特徴分析データを用いて行う。すなわち、雑音の種類は、雑音区間の特徴データ (たとえば  $n 1$  フレーム分の特徴データ) を用いて、雑音が前述した 3 種類の雑音 (雑音  $N 1$ ,  $N 2$ ,  $N 3$ ) のどれに近いかを判定することができ、 $S/N$  比は雑音区間の特徴データから得られるパワーの大きさと、音声区間の特徴データから得られるパワーの大きさによって求めることができる。

#### 【0 1 1 8】

そして、これら雑音の種類判定と  $S/N$  比の大きさに基づいて、どの音響モデルを用いるかの音響モデル選択動作を行う。この音響モデル選択動作は、前述の第 1 および第 2 の実施の形態同様、たとえば、雑音の種類が雑音  $N 1$  に近いと判定され、かつ、 $S/N$  比が  $L 1$  未満であった場合には、音響モデル  $M 1$  が選択されるといった動作である。

#### 【0 1 1 9】

この音響モデル選択処理が終了すると、次に、音声認識を行うに必要な音声特徴データを得るための特徴データ生成処理が CMN 演算部 3 1 によって行われる (ステップ s 4 9, s 5 0)。この特徴データ生成処理は、前述した雑音除去法としての CMN 法を用いて行う。

#### 【0 1 2 0】

この CMN 法は、第 2 の実施の形態で説明したように、音声区間の特徴分析結果による特徴ベクトル ( $n 2$  フレーム分の特徴ベクトル) を用い、その  $n 2$  フレーム分の平均の特徴ベクトルを前述同様の手順で求める (求められた平均の特徴ベクトルを  $C_m$  とする)。この平均の特徴ベクトル  $C_m$  を用い、音声区間 (ここでは 2 0 フレーム分) の特徴ベクトルを再計算する。つまり、 $C 1' = C 1 - C_m$ ,  $C 2' = C 2 - C_m$ ,  $\dots$ ,  $C 2 0' = C 2 0 - C_m$  を行う。

#### 【0 1 2 1】

そして、求められた  $C1'$  ,  $C1'$  ,  $\dots$  ,  $C20'$  が得られた 20 フレーム分のそれぞれのフレームごとの特徴ベクトルとなる。そして、このそれぞれのフレームごとの特徴ベクトル  $C1'$  ,  $C1'$  ,  $\dots$  ,  $C20'$  が音声認識部 10 に与えられ、音声認識部 10 では、選択された音響モデルと言語モデル記憶部 11 に保存されている言語モデルを用いた音声認識処理を行う。

#### 【0122】

このように、第 3 の実施の形態においても前述した第 1 および第 2 の実施の形態と同様、雑音の種類と  $S/N$  比の大きさに応じた音響モデルが選択され、選択された音響モデルと予め用意されている言語モデルを用いて音声認識を行うようにしている。

#### 【0123】

この第 3 の実施の形態では、 $SS$  法（または  $CSS$  法）と  $CMN$  法の両方を用いた音響モデルを作成し、実際の音声認識を行う場合も、 $SS$  法（または  $CSS$  法）を用いた雑音除去を行ったのちに、その雑音除去された音声データに対し、 $CMN$  法で特徴ベクトルを生成し、それを音声認識用の特徴ベクトルとして音声認識部 10 に渡すようにしているので、さらに高い認識性能を得ることができ、また、この第 3 の実施の形態では、加法性雑音や乗法性雑音など幅広い雑音に対する対応が可能となる。

#### 【0124】

なお、本発明は以上説明した実施の形態に限定されるものではなく、本発明の要旨を逸脱しない範囲で種々変形実施可能となるものである。たとえば、前述の各実施の形態では、雑音の種類は、雑音  $N1$  , 雑音  $N2$  , 雑音  $N3$  の 3 種類とし、 $S/N$  比はこれら各雑音について 2 段階の大きさとした例を示したが、これに限られるものではない。

#### 【0125】

特に、雑音の種類は、たとえば、自動車の走行音、エアコンの運転音、街中の雑踏というようにそれぞれを単独の雑音として考えるのではなく、幾つかの雑音を組み合わせたものを 1 つの雑音として考えるようにしてもよい。

#### 【0126】

一例として、雑音のない環境下で収録した音声データに、自動車の走行音とエアコンの運転音を同時に重畳させた音声データを生成し、この生成された音声データから所定の雑音除去方法を用いて雑音を除去したのちに、その雑音の除去された音声データを用いて学習した音声認識用の音響モデルを作成しておくこともできる。

## 【 0 1 2 7 】

このように、機器の使用される環境下に存在しやすい定常雑音を組み合わせて作成された音響モデルを任意に複数種類作成することが可能であるので、個々の機器対応に最適な幾つかの音響モデルを用意しておくことで、より一層、高い認識率を得ることができる。さらに、これらそれぞれの雑音について、 $S/N$ 比の異なるものを作成しておけば、より好結果が得られる。

## 【 0 1 2 8 】

また、図 1、図 3、図 5 で示された音声認識装置の構成は、それぞれ実施の形態の例を示すもので、これらの図で示した通りに構成する必要はない。たとえば、雑音種類を判定する手段と音響モデルを選択する手段を、雑音種類判定手段／音響モデル選択手段 6 として 1 つにまとめたものとしたが、雑音種類判定手段と音響モデル選択手段というようにそれぞれを別個の構成要素として設けるようにしてもよいことは勿論である。

## 【 0 1 2 9 】

さらに、前述の各実施の形態では、種類の異なる複数（3 種類）の雑音を用意し、それぞれの雑音について複数段階（2 段階）の  $S/N$  比を設定した例を説明したが、本発明は、ある特定の決まった雑音（1 種類の雑音）について複数の  $S/N$  を考慮した音声認識を行うことも可能である。

## 【 0 1 3 0 】

その場合、 $S/N$  比の異なるある特定の種類の雑音がそれぞれの  $S/N$  比ごとに重畳されたそれぞれの音声データを作成し、これらそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルを用いることで、それぞれの  $S/N$  比に対応する音響モデル群を作成しておく。

## 【 0 1 3 1 】

そして、実際の認識時には、雑音が重畳された認識対象音声データに対し、重畳されている  $S/N$  比を判定し、その判定結果に基づいて、それぞれの  $S/N$  比に対応した音響モデル群の中から所定の音響モデルを選択し、前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行い、その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行うようにしている。

## 【 0 1 3 2 】

その場合の音声認識装置は、ここでは図示しないが、 $S/N$  比の異なる雑音が雑音の種類ごとに重畳されたそれぞれの音声データを作成し、これらそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データの特徴ベクトルによって作成された  $S/N$  比に対応する音響モデル群と、この音響モデル群を記憶する音響モデル群記憶手段と、雑音が重畳された認識対象音声データに対し、重畳されている雑音の  $S/N$  比を判定する  $S/N$  比判定手段と、その判定結果に基づいて、前記  $S/N$  比に対応した音響モデル群の中から所定の音響モデルを選択する音響モデル群選択手段と、前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行う雑音除去手段と、その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行う音声認識手段とを有した構成とする。

## 【 0 1 3 3 】

なお、この場合も雑音除去手法としては、 $SS$  法（または  $CSS$  法）や  $CMN$  法を用いることが可能で、第 1 の実施の形態、第 2 の実施の形態、さらには、第 3 の実施の形態で説明した処理に準じた処理を行うことで、雑音が重畳された認識対象音声データから  $S/N$  比の大きさを判定し、 $S/N$  比の大きさに応じた音響モデルが選択され、その選択された音響モデルを用いて音声認識を行うことができる。

## 【 0 1 3 4 】

これは、雑音の種類は特定できても、その大きさ（ $S/N$  比）が変動すること

の多い環境下での音声認識を行う場合に好都合なものとなり、そのような環境下での認識率を高くすることができる。この場合、雑音の種類は特定されていることから、雑音の種類を判定する必要がないので、全体の演算量を少なくすることができ、演算能力のより低いCPUでも十分対応できるものとなる。

## 【0135】

また、前述の各実施の形態では、雑音除去手法として、SS法（またはCSS法）やCMN法を用いた例について説明したが、これらSS法（またはCSS法）やCMN法そのものでなく、それらをベースとしてそれらを変形した方法（たとえば、CMN法には、非音声区と音声区間を区別してCMNを行う方法もある）であってもよい。

## 【0136】

また、音声特徴ベクトルとしては、 $\Delta$ ケプストラム係数や $\Delta$ パワーなどを用いてもよい。

## 【0137】

また、本発明は、以上説明した本発明を実現するための処理手順が記述された処理プログラムを作成し、その処理プログラムをフロッピーディスク、光ディスク、ハードディスクなどの記録媒体に記録しておくことができ、本発明はその処理プログラムが記録された記録媒体をも含むものである。また、ネットワークから当該処理プログラムを得るようにしてもよい。

## 【0138】

## 【発明の効果】

以上説明したように本発明は、種類の異なる雑音が重畳されたそれぞれの音声データを作成し、これらそれぞれの雑音が重畳されたそれぞれの音声データに対し、所定の雑音除去手法を用いて雑音除去を行い、その雑音除去されたそれぞれの音声データを用いて、雑音の種類に対応する音響モデルを作成しておく。そして、実際の認識時には、雑音が重畳された認識対象音声データに対し、重畳されている雑音の種類を判定して、その判定結果に基づいて、前記雑音の種類に対応した音響モデルの中から所定の音響モデルを選択するとともに、前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去

を行い、その雑音除去された音声データに対し、前記選択された音響モデルを用いて音声認識を行うようにしている。

## 【 0 1 3 9 】

これによって、重畳されている雑音の種類に応じた最適な音響モデルを用いての認識処理が可能となり、所定の雑音の存在する環境下であっても高い認識率を得ることができる。

## 【 0 1 4 0 】

特に、機器の使用環境に 2、3 種類の雑音が定常的に存在するような場合、これらの雑音ごとの音響モデルを作成し、その音響モデルを用いて、上述したような音声認識処理を行うことで、高い認識率を実現できる。

## 【 0 1 4 1 】

そして、本発明で用いられる雑音除去手法の 1 つとしては、スペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法による雑音除去手法であって、その場合、前記音響モデル作成時における雑音除去は、スペクトラル・サブストラクション法または連続スペクトラル・サブトラクション法を用いて行う。また、実際の認識時には、雑音区間の特徴分析データによって、重畳されている雑音の種類を判定したのち、その判定結果に基づいて、最適な音響モデルを選択するとともに、前記雑音が重畳された認識対象音声データに対し、スペクトラム・サブストラクション法による雑音除去手法を用いて雑音除去を行い、その雑音除去された音声データを特徴分析して得られた結果に対し、前記選択された音響モデルを用いて音声認識を行うようにしている。

## 【 0 1 4 2 】

このように、雑音除去方法としてスペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法を用いることにより、雑音除去処理を少ない演算量で行うことができ、演算能力の低い CPU でも十分対応することができる。これにより、小規模で安価なハードウェア上での実現が可能となる。また、このスペクトラル・サブトラクション法または連続スペクトラル・サブトラクション法は自動車の走行音やエアコンの運転音、街中の雑踏などの雑音（一般に加法性雑音といわれている）の除去に効果があるとされているので、このような雑

音の多い環境下で用いられることが多い機器に適用されることで大きな効果が得られる。

## 【 0 1 4 3 】

また、雑音除去手法の他の例として、ケプストラム平均正規化法による雑音除去手法を用いることもこともできる。その場合、前記音響モデル作成時における雑音除去は、ケプストラム平均正規化法を用いて行う。また、実際の認識時には、雑音区間の特徴分析データによって、重畳されている雑音の種類を判定したのち、その判定結果に基づいて、最適な音響モデルを選択するとともに、前記雑音が重畳された認識対象音声データの音声区間に対し、ケプストラム平均正規化法を用いて雑音除去処理を行い、その雑音除去処理によって得られた特徴ベクトルに対し、前記選択された音響モデルを用いて音声認識を行うようにしている。

## 【 0 1 4 4 】

このように、雑音除去方法としてケプストラム平均正規化法を用いることにより、上述同様、雑音除去処理を少ない演算量で行うことができ、演算能力の低いCPUでも十分対応することができる。

## 【 0 1 4 5 】

これにより、小規模で安価なハードウェア上での実現が可能となる。また、このケプストラム平均正規化法はマイクロホンの特性やエコーなど空間伝達特性に由来する歪みなどの雑音（一般に乗法性雑音といわれている）の除去に効果があるとされているので、このような雑音が発生しやすい環境下で用いられることが多い機器に適用されることで大きな効果が得られる。

## 【 0 1 4 6 】

さらに、それぞれの雑音対応の音響モデルは、雑音の種類に加え、それぞれの雑音の種類ごとに複数段階のS/N比にも対応した音響モデルとし、実際の認識時には、前記雑音が重畳された認識対象音声データに対し、雑音区間の雑音の大きさと音声区間の音声の大きさからS/N比を求め、求められたS/N比と雑音の種類に応じた音響モデルを選択するようにしているので、雑音の種類だけではなくその大きさに応じた最適な音響モデルを用いての認識が行える。これによって、それぞれの雑音環境下において音声認識を行う際、より一層、高い認識率を

得ることが可能となる。

【0147】

また、スペクトラル・サブストラクション法または連続スペクトラル・サブストラクション法とケプストラム平均正規化法の両方を用いた音響モデルを作成することも可能である。この場合、実際の音声認識を行う場合も、スペクトラル・サブストラクション法または連続スペクトラル・サブストラクション法を用いた雑音除去を行ったのちに、その雑音除去された音声データに対し、ケプストラム平均正規化法で特徴ベクトルを生成し、それを音声認識用の特徴ベクトルとして音声認識部に渡すようにしているので、さらに高い認識性能を得ることができ、また、この場合、前述した加法性雑音や乗法性雑音など幅広い雑音に対する対応が可能となる。

【0148】

さらに、ある特定の決まった雑音について複数の $S/N$ に対応する音響モデル群を作成しておき、実際の認識時には、雑音が重畳された認識対象音声データに対し、重畳されている $S/N$ 比を判定し、その判定結果に基づいて、それぞれの $S/N$ 比に対応した音響モデル群の中から所定の音響モデルを選択し、前記雑音が重畳された認識対象音声データに対し、前記所定の雑音除去方法を用いて雑音除去を行い、その雑音除去された音声データの特徴ベクトルに対し前記選択された音響モデルを用いて音声認識を行うようにすることもできる。

【0149】

これによれば、雑音の種類は決まってもその大きさ（ $S/N$ 比）が変動することの多い環境下での音声認識を行う場合に好都合なものとなり、そのような環境下での認識率を高くすることができる。この場合、雑音の種類は特定されているので、雑音の種類を判定する必要がなく、演算量を少なくすることができ、演算能力がより低いCPUでも十分対応できるものとなる。

【図面の簡単な説明】

【図1】

本発明の音声認識装置の第1の実施の形態を説明するための構成図である。

【図2】



第 1 の実施の形態の処理手順を説明するためのフローチャートである。

【図 3】

本発明の音声認識装置の第 2 の実施の形態を説明するための構成図である。

【図 4】

第 2 の実施の形態の処理手順を説明するためのフローチャートである。

【図 5】

本発明の音声認識装置の第 3 の実施の形態を説明するための構成図である。

【図 6】

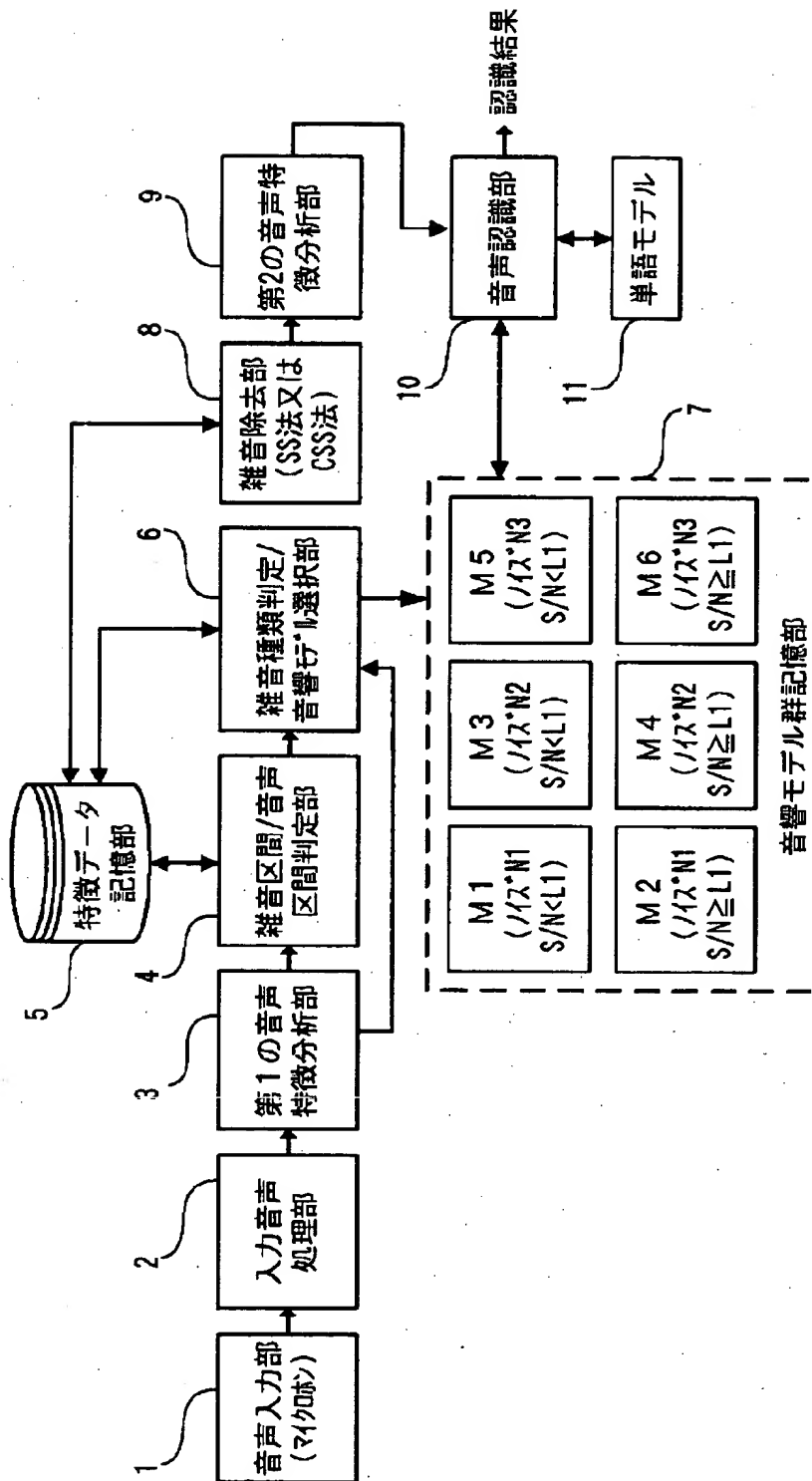
第 3 の実施の形態の処理手順を説明するためのフローチャートである。

【符号の説明】

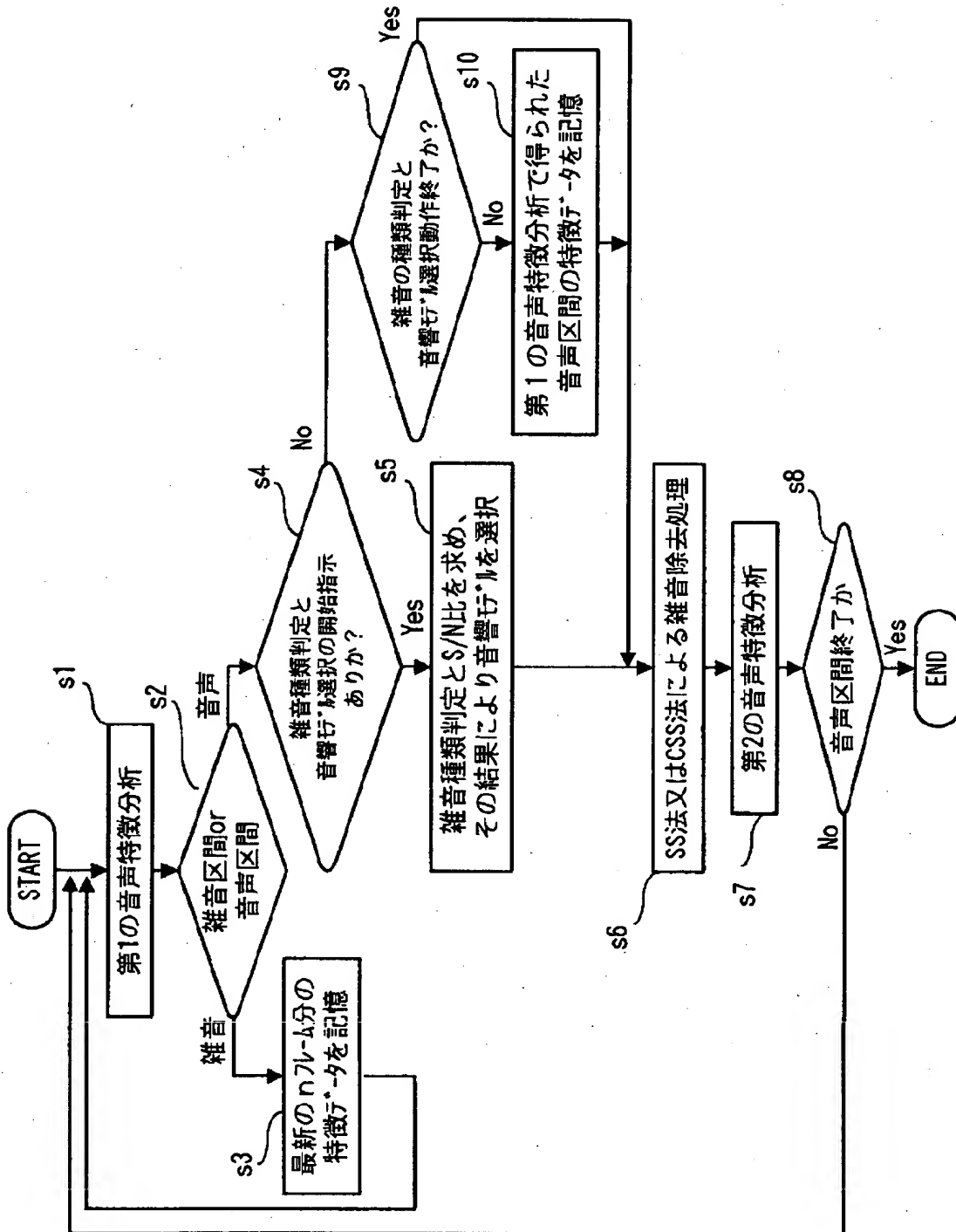
- 1 音声入力部
- 2 入力音声処理部
- 3 第 1 の音声特徴分析部
- 4 雑音区間／音声区間判定部
- 5 特徴データ記憶部
- 6 雑音種類判定／音響モデル選択部
- 7 音響モデル群記憶部
- 8 雑音除去部
- 9 第 2 の音声特徴分析部
- 10 音声認識部
- 11 言語モデル記憶部
- 21 特徴分析部
- 31 CNN 演算部（CNN 法による雑音除去部）

【書類名】 図面

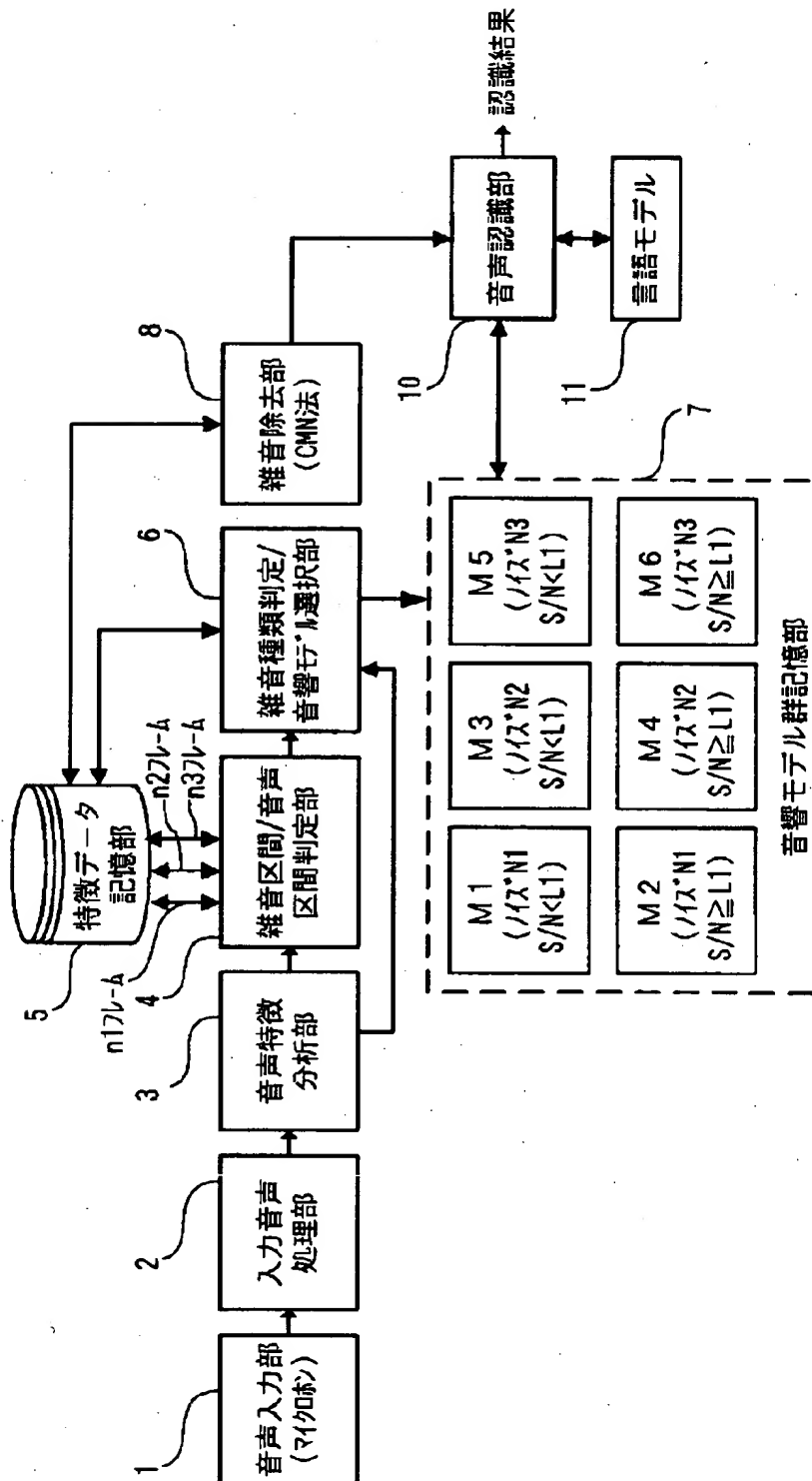
【図 1】



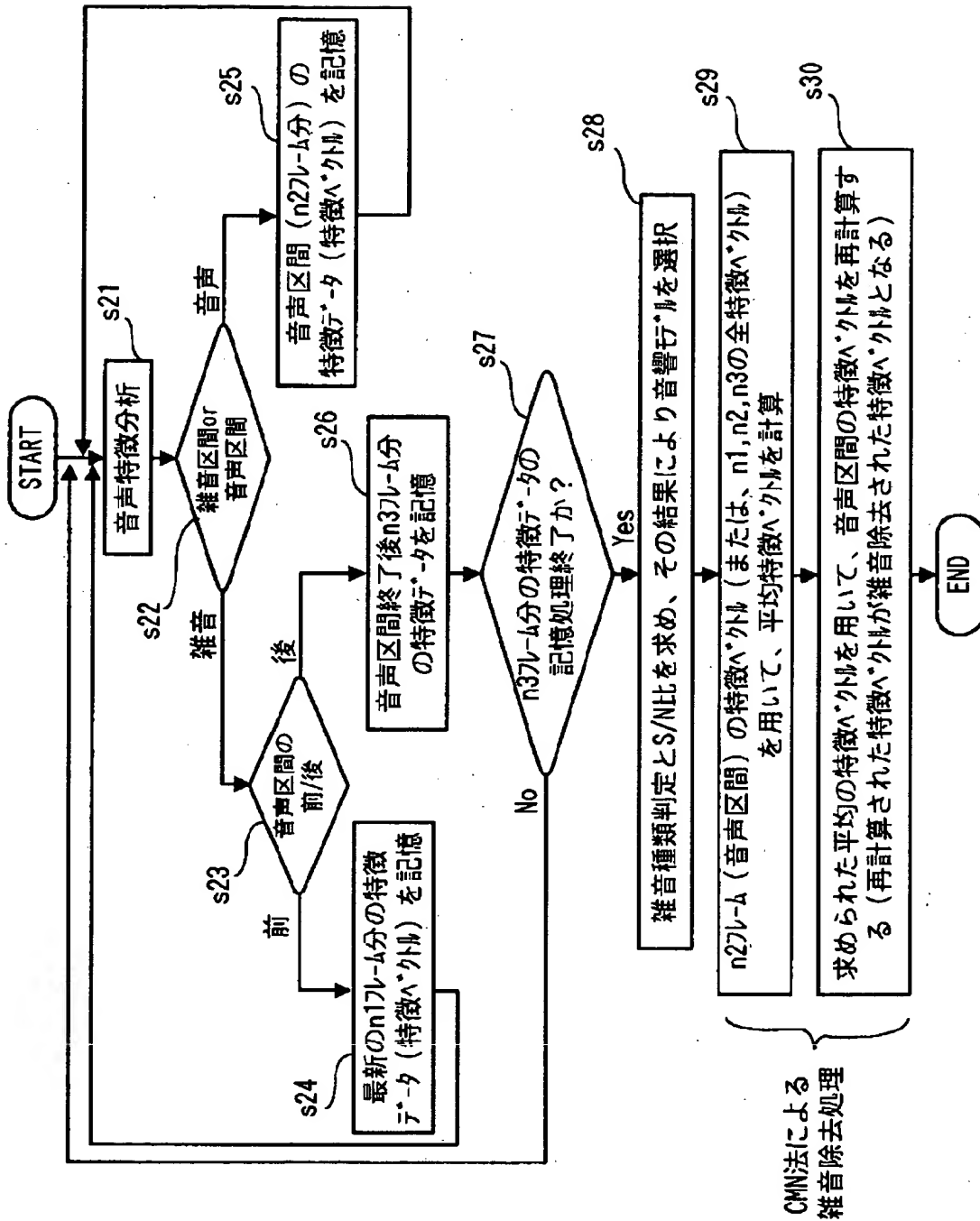
【図 2】



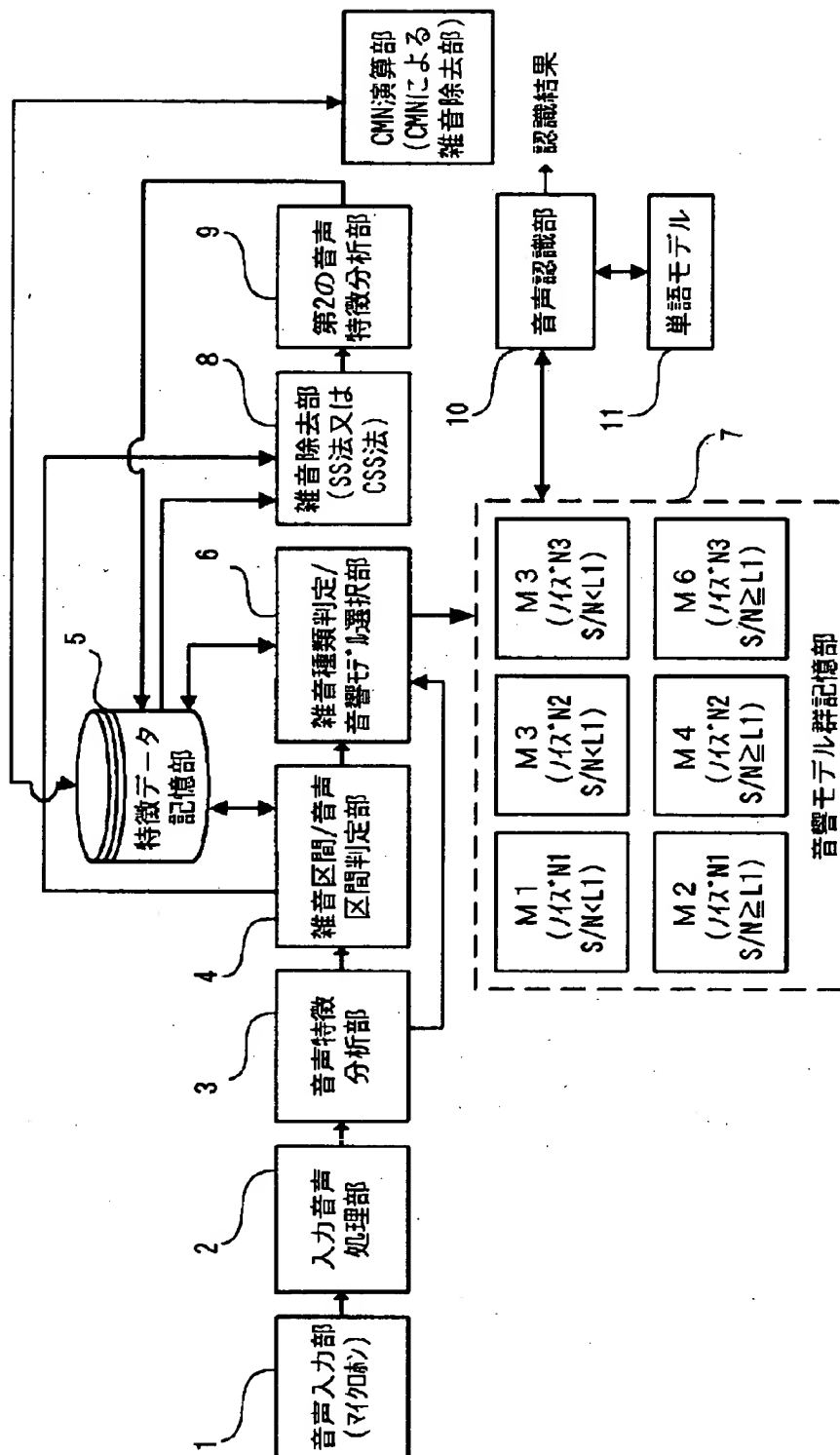
【図 3】



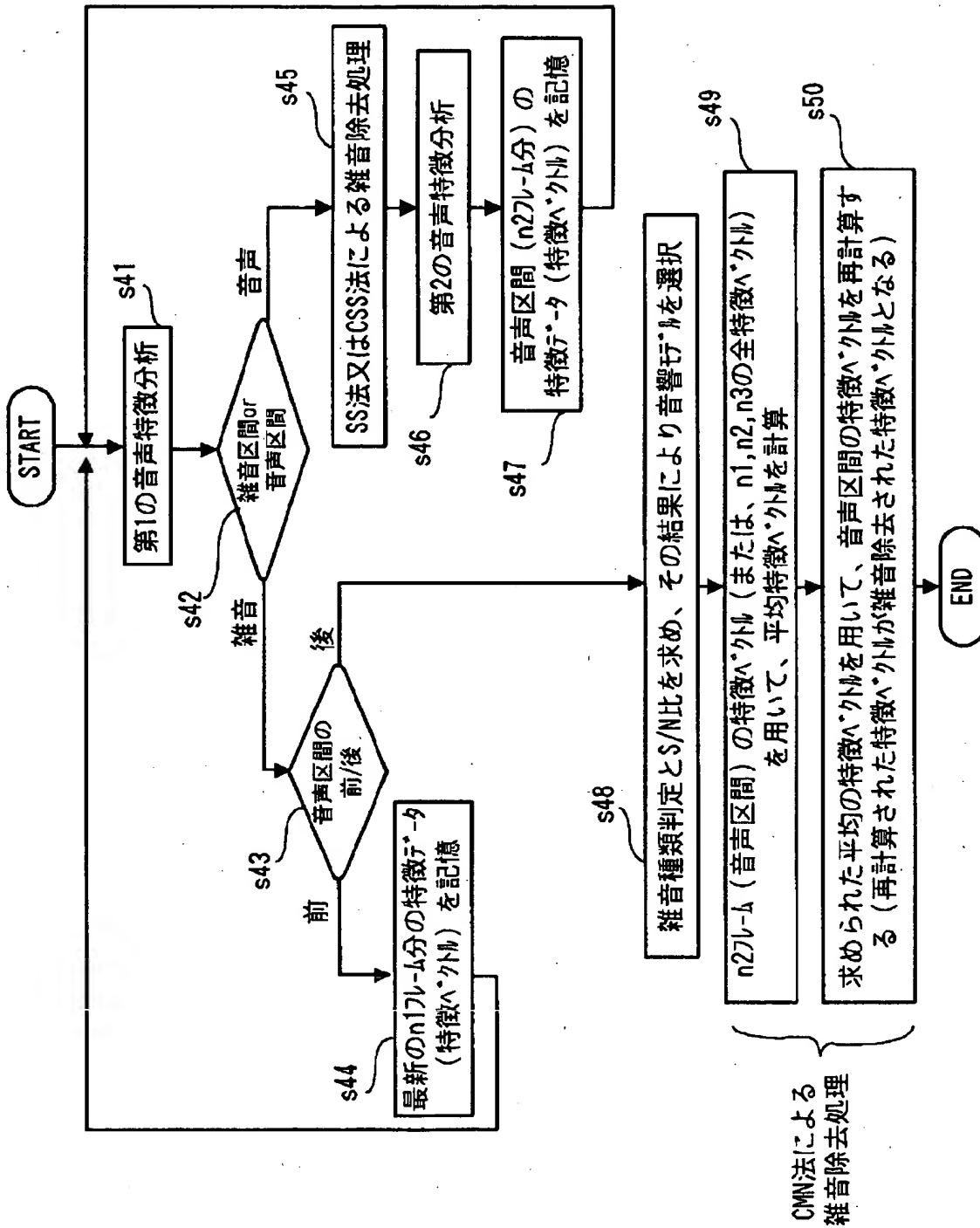
【図4】



【图 5】



【図6】



【書類名】 要約書

【要約】

【課題】複数種の雑音の存在する環境下であっても高い認識率を得るようにする

【解決手段】種類の異なる雑音が重畳されたそれぞれの音声データに対し、スペクトラル・サブトラクションの雑音除去手法（ＳＳ法という）を用いて雑音除去を行い、雑音除去後のそれぞれの音声データを特徴分析して得られた特徴ベクトルにより作成された雑音の種類対応の音響モデルを用意する。認識時には、認識対象音声データに対し、第１の音声特徴分析を行い、雑音区間／音声区間を判定し、雑音区間の場合にはその特徴データを保存し（ステップｓ１～ｓ３）、音声区間の場合には保存された特徴データによって雑音種類を判定し、その結果によって、対応する音響モデルを選択する（ステップｓ５）。そして、認識対象音声データに対し、ＳＳ法による雑音除去を行い、その雑音除去後の音声データに対し、音声認識用の特徴ベクトルを得るための第２の音声特徴分析を行う（ステップｓ６，ｓ７）。

【選択図】 図２



出 願 人 履 歴 情 報

識別番号 [000002369]

1. 変更年月日 1990年 8月20日

[変更理由] 新規登録

住 所 東京都新宿区西新宿2丁目4番1号

氏 名 セイコーエプソン株式会社